ARTIFICIAL ANTIBODY POLYPEPTIDES

Portions of the present invention were made with support of the United States Government via a grant from the National Institutes of Health under grant number GM 55042 The U.S. Government therefore may have certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates generally to the field of the production and selection of binding and catalytic polypeptides by the methods of molecular biology. The invention specifically relates to the generation of both nucleic acid and polypeptide libraries encoding the molecular scaffolding of a modified Fibronectin Type III (Fn3) molecule. The invention also relates to "artificial mini-antibodies" or "monobodies," *i.e.*, polypeptides containing an Fn3 scaffold onto which loop regions capable of binding to a variety of different molecular structures (such as antibody binding sites) have been grafted.

BACKGROUND OF THE INVENTION

Antibody structure

A standard antibody (Ab) is a tetrameric structure consisting of two identical immunoglobulin (Ig) heavy chains and two identical light chains. The heavy and light chains of an Ab consist of different domains. Each light chain has one variable domain (VL) and one constant domain (CL), while each heavy chain has one variable domain (VH) and three or four constant domains (CH) (Alzari *et al.*, 1988). Each domain, consisting of ~110 amino acid residues, is folded into a characteristic β-sandwich structure formed from two β-sheets packed against each other, the immunoglobulin fold. The VH and VL domains each have three complementarity determining regions (CDR1-3) that are loops, or turns, connecting β-strands at one end of the domains (Fig. 1: A, C). The variable regions of both the light and heavy chains generally contribute to

antigen specificity, although the contribution of the individual chains to specificity is not always equal. Antibody molecules have evolved to bind to a large number of molecules by using six randomized loops (CDRs). However, the size of the antibodies and the complexity of six loops represents a major design hurdle if the end result is to be a relatively small peptide ligand.

Antibody substructures

Functional substructures of Abs can be prepared by proteolysis and by recombinant methods. They include the Fab fragment, which contains the VH-CH1 domains of the heavy chain and the VL-CL1 domains of the light chain joined by a single interchain disulfide bond, and the Fv fragment, which contains only the VH and VL domains. In some cases, a single VH domain retains significant affinity (Ward *et al.*, 1989). It has also been shown that a certain monomeric κ light chain will specifically bind to its cognate antigen. (L. Masat *et al.*, 1994). Separated light or heavy chains have sometimes been found to retain some antigen-binding activity (Ward *et al.*, 1989). These antibody fragments are not suitable for structural analysis using NMR spectroscopy due to their size, low solubility or low conformational stability.

Another functional substructure is a single chain Fv (scFv), made of the variable regions of the immunoglobulin heavy and light chain, covalently connected by a peptide linker (S-z Hu *et al.*, 1996). These small (M_r 25,000) proteins generally retain specificity and affinity for antigen in a single polypeptide and can provide a convenient building block for larger, antigen-specific molecules. Several groups have reported biodistribution studies in xenografted athymic mice using scFv reactive against a variety of tumor antigens, in which specific tumor localization has been observed. However, the short persistence of scFvs in the circulation limits the exposure of tumor cells to the scFvs, placing limits on the level of uptake. As a result, tumor uptake by scFvs in animal studies has generally been only 1-5%ID/g as opposed to intact antibodies that can localize in tumors ad 30-40 %ID/g and have reached levels as high as 60-70 %ID/g.

A small protein scaffold called a "minibody" was designed using a part

of the Ig VH domain as the template (Pessi *et al.*, 1993). Minibodies with high affinity (dissociation constant (K_d) ~10^{-7} M) to interleukin-6 were identified by randomizing loops corresponding to CDR1 and CDR2 of VH and then selecting mutants using the phage display method (Martin *et al.*, 1994). These experiments demonstrated that the essence of the Ab function could be transferred to a smaller system. However, the minibody had inherited the limited solubility of the VH domain (Bianchi *et al.*, 1994).

It has been reported that camels (*Camelus dromedarius*) often lack variable light chain domains when IgG-like material from their serum is analyzed, suggesting that sufficient antibody specificity and affinity can be derived form VH domains (three CDR loops) alone. Davies and Riechmann recently demonstrated that "camelized" VH domains with high affinity (K_d ~ 10^{-7} M) and high specificity can be generated by randomizing only the CDR3. To improve the solubility and suppress nonspecific binding, three mutations were introduced to the framework region (Davies & Riechmann, 1995). It has not been definitively shown, however, that camelization can be used, in general, to improve the solubility and stability of VHs.

An alternative to the "minibody" is the "diabody." Diabodies are small bivalent and bispecific antibody fragments, *i.e.*, they have two antigen-binding sites. The fragments contain a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) on the same polypeptide chain (V_H-V_L). Diabodies are similar in size to an Fab fragment. By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. These dimeric antibody fragments, or "diabodies," are bivalent and bispecific (P. Holliger *et al.*, 1993).

Since the development of the monoclonal antibody technology, a large number of 3D structures of Ab fragments in the complexed and/or free states have been solved by X-ray crystallography (Webster *et al.*, 1994; Wilson & Stanfield, 1994). Analysis of Ab structures has revealed that five out of the six CDRs have limited numbers of peptide backbone conformations, thereby permitting one to predict the backbone conformation of CDRs using the so-

called canonical structures (Lesk & Tramontano, 1992; Rees *et al.*, 1994). The analysis also has revealed that the CDR3 of the VH domain (VH-CDR3) usually has the largest contact surface and that its conformation is too diverse for canonical structures to be defined; VH-CDR3 is also known to have a large variation in length (Wu *et al.*, 1993). Therefore, the structures of crucial regions of the Ab-antigen interface still need to be experimentally determined.

Comparison of crystal structures between the free and complexed states has revealed several types of conformational rearrangements. They include side-chain rearrangements, segmental movements, large rearrangements of VH-CDR3 and changes in the relative position of the VH and VL domains (Wilson & Stanfield, 1993). In the free state, CDRs, in particular those which undergo large conformational changes upon binding, are expected to be flexible. Since X-ray crystallography is not suited for characterizing flexible parts of molecules, structural studies in the solution state have not been possible to provide dynamic pictures of the conformation of antigen-binding sites.

Mimicking the antibody-binding site

CDR peptides and organic CDR mimetics have been made (Dougall *et al.*, 1994). CDR peptides are short, typically cyclic, peptides which correspond to the amino acid sequences of CDR loops of antibodies. CDR loops are responsible for antibody-antigen interactions. Organic CDR mimetics are peptides corresponding to CDR loops which are attached to a scaffold, *e.g.*, a small organic compound.

CDR peptides and organic CDR mimetics have been shown to retain some binding affinity (Smyth & von Itzstein, 1994). However, as expected, they are too small and too flexible to maintain full affinity and specificity. Mouse CDRs have been grafted onto the human Ig framework without the loss of affinity (Jones *et al.*, 1986; Riechmann *et al.*, 1988), though this "humanization" does not solve the above-mentioned problems specific to solution studies.

Mimicking natural selection processes of Abs

In the immune system, specific Abs are selected and amplified from a

large library (affinity maturation). The processes can be reproduced *in vitro* using combinatorial library technologies. The successful display of Ab fragments on the surface of bacteriophage has made it possible to generate and screen a vast number of CDR mutations (McCafferty *et al.*, 1990; Barbas *et al.*, 1991; Winter *et al.*, 1994). An increasing number of Fabs and Fvs (and their derivatives) is produced by this technique, providing a rich source for structural studies. The combinatorial technique can be combined with Ab mimics.

A number of protein domains that could potentially serve as protein scaffolds have been expressed as fusions with phage capsid proteins. Review in Clackson & Wells, Trends Biotechnol. 12:173-184 (1994). Indeed, several of these protein domains have already been used as scaffolds for displaying random peptide sequences, including bovine pancreatic trypsin inhibitor (Roberts *et al.*, PNAS 89:2429-2433 (1992)), human growth hormone (Lowman *et al.*, Biochemistry 30:10832-10838 (1991)), Venturini *et al.*, Protein Peptide Letters 1:70-75 (1994)), and the IgG binding domain of *Streptococcus* (O'Neil *et al.*, Techniques in Protein Chemistry V (Crabb, L,. ed.) pp. 517-524, Academic Press, San Diego (1994)). These scaffolds have displayed a single randomized loop or region.

Researchers have used the small 74 amino acid α-amylase inhibitor Tendamistat as a presentation scaffold on the filamentous phage M13 (McConnell and Hoess, 1995). Tendamistat is a β-sheet protein from *Streptomyces tendae*. It has a number of features that make it an attractive scaffold for peptides, including its small size, stability, and the availability of high resolution NMR and X-ray structural data. Tendamistat's overall topology is similar to that of an immunoglobulin domain, with two β-sheets connected by a series of loops. In contrast to immunoglobulin domains, the β-sheets of Tendamistat are held together with two rather than one disulfide bond, accounting for the considerable stability of the protein. By analogy with the CDR loops found in immunoglobulins, the loops the Tendamistat may serve a similar function and can be easily randomized by in vitro mutagenesis.

Tendamistat, however, is derived from *Streptomyces tendae*. Thus, while Tendamistat may be antigenic in humans, its small size may reduce or

inhibit its antigenicity. Also, Tendamistat's stability is uncertain. Further, the stability that is reported for Tendamistat is attributed to the presence of two disulfide bonds. Disulfide bonds, however, are a significant disadvantage to such molecules in that they can be broken under reducing conditions and must be properly formed in order to have a useful protein structure. Further, the size of the loops in Tendamistat are relatively small, thus limiting the size of the inserts that can be accommodated in the scaffold. Moreover, it is well known that forming correct disulfide bonds in newly synthesized peptides is not straightforward. When a protein is expressed in the cytoplasmic space of *E. coli*, the most common host bacterium for protein overexpression, disulfide bonds are usually not formed, potentially making it difficult to prepare large quantities of engineered molecules.

Thus, there is an on-going need for small, single-chain artificial antibodies for a variety of therapeutic, diagnostic and catalytic applications. In particular, there is an on-going need for artificial antibodies that are structurally stable at neutral pH.

SUMMARY OF THE INVENTION

The present invention provides a fibronectin type III (Fn3) molecule, wherein the Fn3 contains a stabilizing mutation. A stabilizing mutation is defined herein as a modification or change in the amino acid sequence of the Fn3 molecule, such as a substitution of one amino acid for another, that increases the melting point of the molecule by more than 0.1°C as compared to a molecule that is identical except for the change. Alternatively, the change may increase the melting point by more than 0.5°C or even 1.0°C or more. A method for determining the melting point of Fn3 molecules is given in Example 19 below.

The Fn3 may have at least one aspartic acid (Asp) residue and/or at least one glutamic acid (Glu) residue that has been deleted or substituted with at least one other amino acid residue. For example, Asp 7 and/or Asp 23 and/or Glu 9, may have been deleted or substituted with at least one other amino acid residue. Asp 7, Asp 23, or Glu 9, may have been substituted with an asparagine (Asn) or lysine (Lys) residue. The present invention further provides an isolated nucleic

acid molecule and an expression vector encoding an Fn3 molecule wherein the Fn3 contains a stabilizing mutation.

The invention provides a fibronectin type III (Fn3) polypeptide monobody containing a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences wherein the Fn3 contains a stabilizing mutation. One or more of the monobody loop region sequences of the Fn3 polypeptide vary by deletion, insertion or replacement of at least two amino acids from the corresponding loop region sequences in wild-type Fn3. The β-strand domains of the monobody have at least about 50% total amino acid sequence homology to the corresponding amino acid sequence of wild-type Fn3's β-strand domain sequences. Preferably, one or more of the loop regions of the monobody contain amino acid residues:

i) from 15 to 16 inclusive in an AB loop;

ii) from 22 to 30 inclusive in a BC loop;

iii) from 39 to 45 inclusive in a CD loop;

iv) from 51 to 55 inclusive in a DE loop;

v) from 60 to 66 inclusive in an EF loop; and

vi) from 76 to 87 inclusive in an FG loop.

The invention also provides a nucleic acid molecule encoding a Fn3 polypeptide monobody wherein the Fn3 contains a stabilizing mutation, as well as an expression vector containing the nucleic acid molecule and a host cell containing the vector.

The invention further provides a method of preparing a Fn3 polypeptide monobody wherein the Fn3 contains a stabilizing mutation. The method includes providing a DNA sequence encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein at least one loop region of the sequence contains a unique restriction enzyme site. The DNA sequence is cleaved at the unique restriction site. Then a preselected DNA segment is inserted into the restriction site. The preselected DNA segment encodes a peptide capable of binding to a specific binding partner (SBP) or a transition state analog compound (TSAC). The insertion of the preselected DNA segment into the DNA sequence yields a DNA molecule which encodes a

polypeptide monobody having an insertion. The DNA molecule is then expressed so as to yield the polypeptide monobody.

Also provided is a method of preparing a Fn3 polypeptide monobody wherein the Fn3 contains a stabilizing mutation, which method includes providing a replicatable DNA sequence encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein the nucleotide sequence of at least one loop region is known. Polymerase chain reaction (PCR) primers are provided or prepared which are sufficiently complementary to the known loop sequence so as to be hybridizable under PCR conditions, wherein at least one of the primers contains a modified nucleic acid sequence to be inserted into the DNA sequence. PCR is performed using the replicatable DNA sequence and the primers. The reaction product of the PCR is then expressed so as to yield a polypeptide monobody.

The invention provides a further method of preparing a Fn3 polypeptide monobody wherein the Fn3 contains a stabilizing mutation. The method includes providing a replicatable DNA sequence encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein the nucleotide sequence of at least one loop region is known. Site-directed mutagenesis of at least one loop region is performed so as to create an insertion mutation. The resultant DNA including the insertion mutation is then expressed.

Further provided is a variegated nucleic acid library encoding Fn3 polypeptide monobodies including a plurality of nucleic acid species encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein one or more of the monobody loop region sequences vary by deletion, insertion or replacement of at least two amino acids from corresponding loop region sequences in wild-type Fn3, and wherein the β-strand domains of the monobody have at least a 50% total amino acid sequence homology to the corresponding amino acid sequence of β-strand domain sequences of the wild-type Fn3, and wherein the Fn3 contains a stabilizing mutation. The invention also provides a peptide display library derived from the variegated nucleic acid library of the invention. Preferably, the peptide of the

peptide display library is displayed on the surface of a bacteriophage, *e.g.*, a M13 bacteriophage or a fd bacteriophage, or virus.

The invention also provides a method of identifying the amino acid sequence of a polypeptide molecule capable of binding to a specific binding partner (SBP) so as to form a polypeptide:SSP complex, wherein the dissociation constant of the the polypeptide:SBP complex is less than 10^{-6} moles/liter. The method includes the steps of:

a) providing a peptide display library of the invention;

b) contacting the peptide display library of (a) with an immobilized or separable SBP;

c) separating the peptide:SBP complexes from the free peptides;

d) causing the replication of the separated peptides of (c) so as to result in a new peptide display library distinguished from that in (a) by having a lowered diversity and by being enriched in displayed peptides capable of binding the SBP;

e) optionally repeating steps (b), (c), and (d) with the new library of (d); and

f) determining the nucleic acid sequence of the region encoding the displayed peptide of a species from (d) and hence deducing the peptide sequence capable of binding to the SBP.

The present invention also provides a method of preparing a variegated nucleic acid library encoding Fn3 polypeptide monobodies having a plurality of nucleic acid species each including a plurality of loop regions, wherein the species encode a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein one or more of the loop region sequences vary by deletion, insertion or replacement of at least two amino acids from corresponding loop region sequences in wild-type Fn3, and wherein the β-strand domain sequences of the monobody have at least a 50% total amino acid sequence homology to the corresponding amino acid sequences of β-strand domain sequences of the wild-type Fn3, and wherein the Fn3 contains a stabilizing mutation, including the steps of

a) preparing an Fn3 polypeptide monobody having a predetermined

sequence;

b) contacting the polypeptide with a specific binding partner (SBP) so as to form a polypeptide:SSP complex wherein the dissociation constant of the the polypeptide:SBP complex is less than 10^{-6} moles/liter;

c) determining the binding structure of the polypeptide:SBP complex by nuclear magnetic resonance spectroscopy or X-ray crystallography; and

d) preparing the variegated nucleic acid library, wherein the variegation is performed at positions in the nucleic acid sequence which, from the information provided in (c), result in one or more polypeptides with improved binding to the SBP.

Also provided is a method of identifying the amino acid sequence of a polypeptide molecule capable of catalyzing a chemical reaction with a catalyzed rate constant, k_{cat}, and an uncatalyzed rate constant, k_{uncat}, such that the ratio of k_{cat}/k_{uncat} is greater than 10. The method includes the steps of:

a) providing a peptide display library of the invention;

b) contacting the peptide display library of (a) with an immobilized or separable transition state analog compound (TSAC) representing the approximate molecular transition state of the chemical reaction;

c) separating the peptide:TSAC complexes from the free peptides;

d) causing the replication of the separated peptides of (c) so as to result in a new peptide display library distinguished from that in (a) by having a lowered diversity and by being enriched in displayed peptides capable of binding the TSAC;

e) optionally repeating steps (b), (c), and (d) with the new library of (d); and

f) determining the nucleic acid sequence of the region encoding the displayed peptide of a species from (d) and hence deducing the peptide sequence.

The invention also provides a method of preparing a variegated nucleic

acid library encoding Fn3 polypeptide monobodies having a plurality of nucleic acid species each including a plurality of loop regions, wherein the species encode a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein one or more of the loop region sequences vary by deletion, insertion or replacement of at least two amino acids from corresponding loop region sequences in wild-type Fn3, and wherein the β-strand domain sequences of the monobody have at least a 50% total amino acid sequence homology to the corresponding amino acid sequences of β-strand domain sequences of the wild-type Fn3, and wherein the Fn3 contains a stabilizing mutation, including the steps of

a) preparing an Fn3 polypeptide monobody having a predetermined sequence, wherein the polypeptide is capable of catalyzing a chemical reaction with a catalyzed rate constant, k_{cat}, and an uncatalyzed rate constant, k_{uncat}, such that the ratio of k_{cat}/k_{uncat} is greater than 10;

b) contacting the polypeptide with an immobilized or separable transition state analog compound (TSAC) representing the approximate molecular transition state of the chemical reaction;

c) determining the binding structure of the polypeptide:TSAC complex by nuclear magnetic resonance spectroscopy or X-ray crystallography; and

d) preparing the variegated nucleic acid library, wherein the variegation is performed at positions in the nucleic acid sequence which, from the information provided in (c), result in one or more polypeptides with improved binding to or stabilization of the TSAC.

The invention also provides a kit for the performance of any of the methods of the invention. The invention further provides a composition, *e.g.*, a polypeptide, prepared by the use of the kit, or identified by any of the methods of the invention.

The following abbreviations have been used in describing amino acids, peptides, or proteins: Ala or A, Alanine; Arg or R, Arginine; Asn or N

asparagine; Asp or D, aspartic acid; Cys or C, cysteine; Gln or Q, glutamine; Glu or E, glutamic acid; Gly or G, glycine; His or H, histidine; Ile or I, isoleucine; Leu or L, leucine; Lys or K, lysine; Met or M, methionine; Phe or F, phenylalanine; Pro or P, proline; Ser or S, serine; Thr or T, threonine; Trp or W, tryptophan; Tyr or Y, tyrosine; Val or V, valine.

The following abbreviations have been used in describing nucleic acids, DNA, or RNA: A, adenosine; T, thymidine; G, guanosine; C, cytosine.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. β-Strand and loop topology (A, B) and MOLSCRIPT representation (C, D; Kraulis, 1991) of the VH domain of anti-lysozyme immunoglobulin D1.3 (A, C; Bhat *et al.*, 1994) and 10th type III domain of human fibronectin (B, D; Main *et al.*, 1992). The locations of complementarity determining regions (CDRs, hypervariable regions) and the integrin-binding Arg-Gly-Asp (RGD) sequence are indicated.

Figure 2. Amino acid sequence (SEQ ID NO:110) and restriction sites of the synthetic Fn3 gene. The residue numbering is according to Main *et al.* (1992). Restriction enzyme sites designed are shown above the amino acid sequence. β-Strands are denoted by underlines. The N-terminal "mq" sequence has been added for a subsequent cloning into an expression vector. The His•tag (Novagen) fusion protein has an additional sequence, MGSSHHHHHHSSGLVPRGSH (SEQ ID NO:114), preceding the Fn3 sequence shown above.

Figure 3. A, Far UV CD spectra of wild-type Fn3 at 25°C and 90°C. Fn3 (50 μM) was dissolved in sodium acetate (50 mM, pH 4.6). B, thermal denaturation of Fn3 monitored at 215 nm. Temperature was increased at a rate of 1°C/min.

Figure 4. A, Cα trace of the crystal structure of the complex of lysozyme (HEL) and the Fv fragment of the anti-hen egg-white lysozyme (anti-HEL) antibody D1.3 (Bhat *et al.*, 1994). Side chains of the residues 99-102 of VH CDR3, which make contact with HEL, are also shown. B, Contact surface area

for each residue of the D1.3 VH-HEL and VH-VL interactions plotted vs. residue number of D1.3 VH. Surface area and secondary structure were determined using the program DSSP (Kabsh and Sander, 1983). C and D, schematic drawings of the β-sheet structure of the F strand-loop-G strand moieties of D1.3 VH (C) and Fn3 (D). The boxes denote residues in β-strands and ovals those not in strands. The shaded boxes indicate residues of which side chains are significantly buried. The broken lines indicate hydrogen bonds.

Figure 5. Designed Fn3 gene showing DNA (SEQ ID NO:111) and amino acid (SEQ ID NO:112) sequences. The amino acid numbering is according to Main *et al.* (1992). The two loops that were randomized in combinatorial libraries are enclosed in boxes.

Figure 6. Map of plasmid pAS45. Plasmid pAS45 is the expression vector of His•tag-Fn3.

Figure 7. Map of plasmid pAS25. Plasmid pAS25 is the expression vector of Fn3.

Figure 8. Map of plasmid pAS38. pAS38 is a phagmid vector for the surface display of Fn3.

Figure 9. (Ubiquitin-1) Characterization of ligand-specific binding of enriched clones using phage enzyme-linked immunosolvent assay (ELISA). Microtiter plate wells were coated with ubiquitin (1 μg/well; "Ligand (+)) and then blocked with BSA. Phage solution in TBS containing approximately 10^{10} colony forming units (cfu) was added to a well and washed with TBS. Bound phages were detected with anti-phage antibody-POD conjugate (Pharmacia) with Turbo-TMB (Pierce) as a substrate. Absorbance was measured using a Molecular Devices SPECTRAmax 250 microplate spectrophotometer. For a control, wells without the immobilized ligand were used. 2-1 and 2-2 denote enriched clones from Library 2 eluted with free ligand and acid, respectively. 4-1 and 4-2 denote enriched clones from Library 4 eluted with free ligand and acid, respectively.

Figure 10. (Ubiquitin-2) Competition phage ELISA of enriched clones. Phage solutions containing approximately 10^{10} cfu were first incubated with free ubiquitin at 4°C for 1 hour prior to the binding to a ligand-coated well. The

wells were washed and phages detected as described above.

Figure 11. Competition phage ELISA of ubiquitin-binding monobody 411. Experimental conditions are the same as described above for ubiquitin. The ELISA was performed in the presence of free ubiquitin in the binding solution. The experiments were performed with four different preparations of the same clone.

Figure 12. (Fluorescein-1) Phage ELISA of four clones, Plb25.1 (containing SEQ ID NO:115), Plb25.4 (containing SEQ ID NO:116), pLB24.1 (containing SEQ ID NO:117) and pLB24.3 (containing SEQ ID NO:118). Experimental conditions are the same as ubiquitin-1 above.

Figure 13. (Fluorescein-2) Competition ELISA of the four clones. Experimental conditions are the same as ubiquitin-2 above.

Figure 14. ^{1}H, ^{15}N-HSQC spectrum of a fluorescence-binding monobody LB25.5. Approximately 20 μM protein was dissolved in 10 mM sodium acetate buffer (pH 5.0) containing 100 mM sodium chloride. The spectrum was collected at 30°C on a Varian Unity INOVA 600 NMR spectrometer.

Figure 15. Characterization of the binding reaction of Ubi4-Fn3 to the target, ubiquitin. (a) Phage ELISA analysis of binding of Ubi4-Fn3 to ubiquitin. The binding of Ubi4-phages to ubiquitin-coated wells was measured. The control experiment was performed with wells containing no ubiquitin.

(b) Competition phage ELISA of Ubi4-Fn3. Ubi4-Fn3-phages were preincubated with soluble ubiquitin at an indicated concentration, followed by the phage ELISA detection in ubiquitin-coated wells.

(c) Competition phage ELISA testing the specificity of the Ubi4 clone. The Ubi4 phages were preincubated with 250 μg/ml of soluble proteins, followed by phage ELISA as in (b).

(d) ELISA using free proteins.

Figure 16. Equilibrium unfolding curves for Ubi4-Fn3 (closed symbols) and wild-type Fn3 (open symbols). Squares indicate data measured in TBS (Tris HCl buffer (50 mM, pH 7.5) containing NaCl (150 mM)). Circles indicate data measured in Gly HCl buffer (20 mM, pH 3.3) containing NaCl (300 mM). The

curves show the best fit of the transition curve based on the two-state model. Parameters characterizing the transitions are listed in Table 8.

Figure 17. (a) ^{1}H, ^{15}N-HSQC spectrum of [^{15}N]-Ubi4-K Fn3. (b). Difference ($\delta_{wild\text{-}type} - \delta_{Ubi4}$) of ^{1}H (b) and ^{15}N (c) chemical shifts plotted versus residue number. Values for residues 82-84 (shown as filled circles) where Ubi4-K deletions are set to zero. Open circles indicate residues that are mutated in the Ubi4-K protein. The locations of β-strands are indicated with arrows.

Figure 18. (A) Guanidine hydrochloride (GuHCl)-induced denaturation of FNfn10 monitored by Trp fluorescence. The fluorescence emission intensity at 355 nm is shown as a function of GuHCl concentration. The lines show the best fits of the data to the two-state transition model. (B) Stability of FN3 at 4 M GuHCl plotted as a function of pH. (C) pH dependence of the m value.

Figure 19. A two-dimensional H(C)CO spectrum of FNfn10 showing the ^{13}C chemical shift of the carboxyl carbon (vertical axis) and the ^{1}H shift of $^{1}H^{\beta}$ of Asp or $^{1}H^{\gamma}$ of Glu, respectively (horizontal axis). Cross peaks are labeled with their respective residue numbers.

Figure 20. pH-Dependent shifts of the ^{13}C chemical shifts of the carboxyl carbons of Asp and Glu residues in FNfn10. Panel A shows data for Asp 3, 67 and 80, and Glu 38 and 47. The lines are the best fits of the data to the Henderson-Hasselbalch equation with one ionizable group (McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Koerner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. & Withers, S. G. (1996) *Biochemistry 35*, 9958-9966). Panel B shows data for Asp 7 and 23 and Glu 9. The continuous lines show the best fits to the Henderson-Hasselbalch equation with two ionizable groups, while the dashed lines show the best fits to the equation with a single ionizable group.

Figure 21. (A) The amino acid sequence of FNfn10 (SEQ ID NO:121) shown according to its topology (Main, A. L., Harvey, T. S., Baron, M., Boyd, J., & Campbell, I. D. (1992) *Cell 71*, 671-678). Asp and Glu residues are highlighted with gray circles. The thin lines and arrows connecting circles indicate backbone hydrogen bonds. (B) A CPK model of FN3 showing the locations of Asp 7 and 23 and Glu 9.

Figure 22. Thermal denaturation of the wild-type and mutant FNfn10 proteins at pH 7.0 and 2.4 in the presence of 6.3 M urea and 0.1 or 1.0 M NaCl. Change in circular dichroism signal at 227 nm is plotted as a function of temperature. The filled circles show the data in the presence of 1 M NaCl and the open circles are data in the presence of 0.1 M NaCl. The left column shows data taken at pH 2.4 and the right column at pH 7.0. The identity of proteins is indicated in the panels.

Figure 23. GuHCl-induce denaturation of FNfn10 mutants monitored with fluorescence. Fluorescence data was converted to the fraction of unfolded protein according to the two-state transition model (Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999) *Biochemistry 38*, 16419-16423), and plotted as a function of GuHCl.

Figure 24. pH Titration of the carboxyl ^{13}C resonance of Asp and Glu residues in D7N (open circles) and D7K (closed circles) FNfn10. Data for the wild-type (crosses) are also shown for comparison. Residue names are denoted in the individual panels.

DETAILED DESCRIPTION OF THE INVENTION

For the past decade the immune system has been exploited as a rich source of *de novo* catalysts. Catalytic antibodies have been shown to have chemoselectivity, enantioselectivity, large rate accelerations, and even an ability to reroute chemical reactions. In most cases the antibodies have been elicited to transition state analog (TSA) haptens. These TSA haptens are stable, low-molecular weight compounds designed to mimic the structures of the energetically unstable transition state species that briefly (approximate half-life 10^{-13} s) appear along reaction pathways between reactants and products. Anti-TSA antibodies, like natural enzymes, are thought to selectively bind and stabilize transition state, thereby easing the passage of reactants to products. Thus, upon binding, the antibody lowers the energy of the actual transition state and increases the rate of the reaction. These catalysts can be programmed to bind to geometrical and electrostatic features of the transition state so that the

reaction route can be controlled by neutralizing unfavorable charges, overcoming entropic barriers, and dictating stereoelectronic features of the reaction. By this means even reactions that are otherwise highly disfavored have been catalyzed (Janda *et al.* 1997). Further, in many instances catalysts have been made for reactions for which there are no known natural or man-made enzymes.

The success of any combinatorial chemical system in obtaining a particular function depends on the size of the library and the ability to access its members. Most often the antibodies that are made in an animal against a hapten that mimics the transition state of a reaction are first screened for binding to the hapten and then screened again for catalytic activity. An improved method allows for the direct selection for catalysis from antibody libraries in phage, thereby linking chemistry and replication.

A library of antibody fragments can be created on the surface of filamentous phage viruses by adding randomized antibody genes to the gene that encodes the phage's coat protein. Each phage then expresses and displays multiple copies of a single antibody fragment on its surface. Because each phage possesses both the surface-displayed antibody fragment and the DNA that encodes that fragment, and antibody fragment that binds to a target can be identified by amplifying the associated DNA.

Immunochemists use as antigens materials that have as little chemical reactivity as possible. It is almost always the case that one wishes the ultimate antibody to interact with native structures. In reactive immunization the concept is just the opposite. One immunizes with compounds that are highly reactive so that upon binding to the antibody molecule during the induction process, a chemical reaction ensues. Later this same chemical reaction becomes part of the mechanism of the catalytic event. In a certain sense one is immunizing with a chemical reaction rather than a substance *per se*. Reactive immunogens can be considered as analogous to the mechanism-based inhibitors that enzymologists use except that they are used in the inverse way in that, instead of inhibiting a mechanism, they induce a mechanism.

Man-made catalytic antibodies have considerable commercial potential in many different applications. Catalytic antibody-based products have been used

successfully in prototype experiments in therapeutic applications, such as prodrug activation and cocaine inactivation, and in nontherapeutic applications, such as biosensors and organic synthesis.

Catalytic antibodies are theoretically more attractive than noncatalytic antibodies as therapeutic agents because, being catalytic, they may be used in lower doses, and also because their effects are unusually irreversible (for example, peptide bond cleavage rather than binding). In therapy, purified catalytic antibodies could be directly administered to a patient, or alternatively the patient's own catalytic antibody response could be elicited by immunization with an appropriate hapten. Catalytic antibodies also could be used as clinical diagnostic tools or as regioselective or stereoselective catalysts in the synthesis of fine chemicals.

I. Mutation of Fn3 loops and grafting of Ab loops onto Fn3

An ideal scaffold for CDR grafting is highly soluble and stable. It is small enough for structural analysis, yet large enough to accommodate multiple CDRs so as to achieve tight binding and/or high specificity.

A novel strategy to generate an artificial Ab system on the framework of an existing non-Ab protein was developed. An advantage of this approach over the minimization of an Ab scaffold is that one can avoid inheriting the undesired properties of Abs. Fibronectin type III domain (Fn3) was used as the scaffold. Fibronectin is a large protein which plays essential roles in the formation of extracellular matrix and cell-cell interactions; it consists of many repeats of three types (I, II and III) of small domains (Baron *et al.*, 1991). Fn3 itself is the paradigm of a large subfamily (Fn3 family or s-type Ig family) of the immunoglobulin superfamily (IgSF). The Fn3 family includes cell adhesion molecules, cell surface hormone and cytokine receptors, chaperonins, and carbohydrate-binding domains (for reviews, see Bork & Doolittle, 1992; Jones, 1993; Bork *et al.*, 1994; Campbell & Spitzfaden, 1994; Harpez & Chothia, 1994).

Recently, crystallographic studies revealed that the structure of the DNA binding domains of the transcription factor NF-kB is also closely related to the

Fn3 fold (Ghosh *et al.*, 1995; Müller *et al.*, 1995). These proteins are all involved in specific molecular recognition, and in most cases ligand-binding sites are formed by surface loops, suggesting that the Fn3 scaffold is an excellent framework for building specific binding proteins. The 3D structure of Fn3 has been determined by NMR (Main *et al.*, 1992) and by X-ray crystallography (Leahy *et al.*, 1992; Dickinson *et al.*, 1994). The structure is best described as a β-sandwich similar to that of Ab VH domain except that Fn3 has seven β-strands instead of nine (Fig. 1). There are three loops on each end of Fn3; the positions of the BC, DE and FG loops approximately correspond to those of CDR1, 2 and 3 of the VH domain, respectively (Fig. 1 C, D).

Fn3 is small (~ 95 residues), monomeric, soluble and stable. It is one of few members of IgSF that do not have disulfide bonds; VH has an interstrand disulfide bond (Fig. 1 A) and has marginal stability under reducing conditions. Fn3 has been expressed in *E. coli* (Aukhil *et al.*, 1993). In addition, 17 Fn3 domains are present just in human fibronectin, providing important information on conserved residues which are often important for the stability and folding (for sequence alignment, see Main *et al.*, 1992 and Dickinson *et al.*, 1994). From sequence analysis, large variations are seen in the BC and FG loops, suggesting that the loops are not crucial to stability. NMR studies have revealed that the FG loop is highly flexible; the flexibility has been implicated for the specific binding of the 10th Fn3 to $\alpha_5\beta_1$ integrin through the Arg-Gly-Asp (RGD) motif. In the crystal structure of human growth hormone-receptor complex (de Vos *et al.*, 1992), the second Fn3 domain of the receptor interacts with hormone via the FG and BC loops, suggesting it is feasible to build a binding site using the two loops.

The tenth type III module of fibronectin has a fold similar to that of immunoglobulin domains, with seven β strands forming two antiparallel β sheets, which pack against each other (Main *et al.*, 1992). The structure of the type II module consists of seven β strands, which form a sandwich of two antiparallel β sheets, one containing three strands (ABE) and the other four strands (C'CFG) (Williams *et al.*, 1988). The triple-stranded β sheet consists of residues Glu-9-Thr-14 (A), Ser-17-Asp-23 (B), and Thr-56-Ser-60 (E). The

majority of the conserved residues contribute to the hydrophobic core, with the invariant hydrophobic residues Trp-22 and Try-68 lying toward the N-terminal and C-terminal ends of the core, respectively. The β strands are much less flexible and appear to provide a rigid framework upon which functional, flexible loops are built. The topology is similar to that of immunoglobulin C domains.

Gene construction and mutagenesis

A synthetic gene for tenth Fn3 of human fibronectin (Fig. 2) was designed which includes convenient restriction sites for ease of mutagenesis and uses specific codons for high-level protein expression (Gribskov *et al.*, 1984).

The gene was assembled as follows: (1) the gene sequence was divided into five parts with boundaries at designed restriction sites (Fig.2); (2) for each part, a pair of oligonucleotides that code opposite strands and have complementary overlaps of ~ 15 bases was synthesized; (3) the two oligonucleotides were annealed and single strand regions were filled in using the Klenow fragment of DNA polymerase; (4) the double-stranded oligonucleotide was cloned into the pET3a vector (Novagen) using restriction enzyme sites at the termini of the fragment and its sequence was confirmed by an Applied Biosystems DNA sequencer using the dideoxy termination protocol provided by the manufacturer; (5) steps 2-4 were repeated to obtain the whole gene (plasmid pAS25) (Fig. 7).

Although the present method takes more time to assemble a gene than the one-step polymerase chain reaction (PCR) method (Sandhu *et al.*, 1992), no mutations occurred in the gene. Mutations would likely have been introduced by the low fidelity replication by Taq polymerase and would have required time-consuming gene editing. The gene was also cloned into the pET15b (Novagen) vector (pEW1). Both vectors expressed the Fn3 gene under the control of bacteriophage T7 promoter (Studler *et al.* 1990); pAS25 expressed the 96-residue Fn3 protein only, while pEW1 expressed Fn3 as a fusion protein with poly-histidine peptide (His•tag). Recombinant DNA manipulations were performed according to Molecular Cloning (Sambrook *et al.*, 1989), unless otherwise stated.

Mutations were introduced to the Fn3 gene using either cassette mutagenesis or oligonucleotide site-directed mutagenesis techniques (Deng & Nickoloff, 1992). Cassette mutagenesis was performed using the same protocol for gene construction described above; double-stranded DNA fragment coding a new sequence was cloned into an expression vector (pAS25 and/or pEW1). Many mutations can be made by combining a newly synthesized strand (coding mutations) and an oligonucleotide used for the gene synthesis. The resulting genes were sequenced to confirm that the designed mutations and no other mutations were introduced by mutagenesis reactions.

Design and synthesis of Fn3 mutants with antibody CDRs

Two candidate loops (FG and BC) were identified for grafting. Antibodies with known crystal structures were examined in order to identify candidates for the sources of loops to be grafted onto Fn3. Anti-hen egg lysozyme (HEL) antibody D1.3 (Bhat *et al.*, 1994) was chosen as the source of a CDR loop. The reasons for this choice were: (1) high resolution crystal structures of the free and complexed states are available (Fig. 4 A; Bhat *et al.*, 1994), (2) thermodynamics data for the binding reaction are available (Tello *et al.*, 1993), (3) D1.3 has been used as a paradigm for Ab structural analysis and Ab engineering (Verhoeyen *et al.*, 1988; McCafferty *et al.*, 1990) (4) site-directed mutagenesis experiments have shown that CDR3 of the heavy chain (VH-CDR3) makes a larger contribution to the affinity than the other CDRs (Hawkins *et al.*, 1993), and (5) a binding assay can be easily performed. The objective for this trial was to graft VH-CDR3 of D1.3 onto the Fn3 scaffold without significant loss of stability.

An analysis of the D1.3 structure (Fig. 4) revealed that only residues 99-102 ("RDYR") (SEQ ID NO:120) make direct contact with hen egg-white lysozyme (HEL) (Fig. 4 B), although VH-CDR3 is defined as longer (Bhat *et al.*, 1994). It should be noted that the C-terminal half of VH-CDR3 (residues 101-104) made significant contact with the VL domain (Fig. 4 B). It has also become clear that D1.3 VH-CDR3 (Fig. 4 C) has a shorter turn between the strands F and G than the FG loop of Fn3 (Fig. 4 D). Therefore, mutant

sequences were designed by using the RDYR (99-102) (SEQ ID NO:120) of D1.3 as the core and made different boundaries and loop lengths (Table 1). Shorter loops may mimic the D1.3 CDR3 conformation better, thereby yielding higher affinity, but they may also significantly reduce stability by removing wild-type interactions of Fn3.

Table 1. Amino acid sequences of D1.3 VH CDR3, VH8 CDR3 and Fn3 FG loop and list of planned mutants.

	96 100 105 • • •	
D1.3	A R E R D Y R L D Y W G Q G	(SEQ ID NO:1)
VH8	A R G A V V S Y Y A M D Y W G Q G	(SEQ ID NO:2)
	75 80 85 • • •	
Fn3	Y A V T G R G D S P A S S K P I	(SEQ ID NO:3)
Mutant	Sequence	
D1.3-1	Y A **E R D Y R L D Y** – – – – P I	(SEQ ID NO:4)
D1.3-2	Y A V **R D Y R L D Y** – – – – P I	(SEQ ID NO:5)
D1.3-3	Y A V **R D Y R L D Y** A S S K P I	(SEQ ID NO:6)
D1.3-4	Y A V **R D Y R L D Y** – – – K P I	(SEQ ID NO:7)
D1.3-5	Y A V **R D Y R** – – – – – S K P I	(SEQ ID NO:8)
D 1.3-6	Y A V T **R D Y R L** – – S S K P I	(SEQ ID NO:9)
D1.3-7	Y A V T **E R D Y R L** – S S K P I	(SEQ ID NO:10)
VH8-1	Y A V **A V V S Y Y A M D Y** – P I	(SEQ ID NO:11)
VH8-2	Y A V T **A V V S Y Y** A S S K P I	(SEQ ID NO:12)

Underlines indicate residues in β-strands. Bold characters indicate replaced residues.

In addition, an anti-HEL single VH domain termed VH8 (Ward *et al.*, 1989) was chosen as a template. VH8 was selected by library screening and, in spite of the lack of the VL domain, VH8 has an affinity for HEL of 27 nM, probably due to its longer VH-CDR3 (Table 1). Therefore, its VH-CDR3 was grafted onto Fn3. Longer loops may be advantageous on the Fn3 framework because they may provide higher affinity and also are close to the loop length of wild-type Fn3. The 3D structure of VH8 was not known and thus the VH8 CDR3 sequence was aligned with that of D1.3 VH-CDR3; two loops were designed (Table 1).

Mutant construction and production

Site-directed mutagenesis experiments were performed to obtain designed sequences. Two mutant Fn3s, D1.3-1 and D1.3-4 (Table 1) were obtained and both were expressed as soluble His•tag fusion proteins. D1.3-4 was purified and the His•tag portion was removed by thrombin cleavage. D1.3-4 is soluble up to at least 1 mM at pH 7.2. No aggregation of the protein has been observed during sample preparation and NMR data acquisition.

Protein expression and purification

E. coli BL21 (DE3) (Novagen) were transformed with an expression vector (pAS25, pEW1 and their derivatives) containing a gene for the wild-type or a mutant. Cells were grown in M9 minimal medium and M9 medium supplemented with Bactotrypton (Difco) containing ampicillin (200 μg/ml). For isotopic labeling, ^{15}N NH_4Cl and/or ^{13}C glucose replaced unlabeled components. 500 ml medium in a 2 liter baffle flask were inoculated with 10 ml of overnight culture and agitated at 37°C. Isopropylthio-β-galactoside (IPTG) was added at a final concentration of 1 mM to initiate protein expression when OD (600 nm) reaches one. The cells were harvested by centrifugation 3 hours after the addition of IPTG and kept frozen at -70°C until used.

Fn3 without His•tag was purified as follows. Cells were suspended in 5 ml/(g cell) of Tris (50 mM, pH 7.6) containing ethylenediaminetetraacetic acid (EDTA; 1 mM) and phenylmethylsulfonyl fluoride (1 mM). HEL was added to a final concentration of 0.5 mg/ml. After incubating the solution for 30 minutes at 37°C, it was sonicated three times for 30 seconds on ice. Cell debris was removed by centrifugation. Ammonium sulfate was added to the solution and precipitate recovered by centrifugation. The pellet was dissolved in 5-10 ml sodium acetate (50 mM, pH 4.6) and insoluble material was removed by centrifugation. The solution was applied to a Sephacryl S100HR column (Pharmacia) equilibrated in the sodium acetate buffer. Fractions containing Fn3 then was applied to a ResourceS column (Pharmacia) equilibrated in sodium acetate (50 mM, pH 4.6) and eluted with a linear gradient of sodium chloride (0-0.5 M). The protocol can be adjusted to purify mutant proteins with different

surface charge properties.

Fn3 with His•tag was purified as follows. The soluble fraction was prepared as described above, except that sodium phosphate buffer (50 mM, pH 7.6) containing sodium chloride (100 mM) replaced the Tris buffer. The solution was applied to a Hi-Trap chelating column (Pharmacia) preloaded with nickel and equilibrated in the phosphate buffer. After washing the column with the buffer, His•tag-Fn3 was eluted in the phosphate buffer containing 50 mM EDTA. Fractions containing His•tag-Fn3 were pooled and applied to a Sephacryl S100-HR column, yielding highly pure protein. The His•tag portion was cleaved off by treating the fusion protein with thrombin using the protocol supplied by Novagen. Fn3 was separated from the His•tag peptide and thrombin by a ResourceS column using the protocol above.

The wild-type and two mutant proteins so far examined are expressed as soluble proteins. In the case that a mutant is expressed as inclusion bodies (insoluble aggregate), it is first examined if it can be expressed as a soluble protein at lower temperature (*e.g.*, 25-30°C). If this is not possible, the inclusion bodies are collected by low-speed centrifugation following cell lysis as described above. The pellet is washed with buffer, sonicated and centrifuged. The inclusion bodies are solubilized in phosphate buffer (50 mM, pH 7.6) containing guanidinium chloride (GdnCl, 6 M) and will be loaded on a Hi-Trap chelating column. The protein is eluted with the buffer containing GdnCl and 50 mM EDTA.

Conformation of mutant Fn3, D1.3-4

The ^{1}H NMR spectra of His•tag D1.3-4 fusion protein closely resembled that of the wild-type, suggesting the mutant is folded in a similar conformation to that of the wild-type. The spectrum of D1.3-4 after the removal of the His•tag peptide showed a large spectral dispersion. A large dispersion of amide protons (7-9.5 ppm) and a large number of downfield (5.0-6.5 ppm) C^{α} protons are characteristic of a β-sheet protein (Wüthrich, 1986).

The 2D NOESY spectrum of D1.3-4 provided further evidence for a preserved conformation. The region in the spectrum showed interactions

between upfield methyl protons (< 0.5 ppm) and methyl-methylene protons. The Val72 γ methyl resonances were well separated in the wild-type spectrum (-0.07 and 0.37 ppm; (Baron *et al.*, 1992)). Resonances corresponding to the two methyl protons are present in the D1.3-4 spectrum (-0.07 and 0.44 ppm). The cross peak between these two resonances and other conserved cross peaks indicate that the two resonances in the D1.3-4 spectrum are highly likely those of Val72 and that other methyl protons are in nearly identical environment to that of wild-type Fn3. Minor differences between the two spectra are presumably due to small structural perturbation due to the mutations. Val72 is on the F strand, where it forms a part of the central hydrophobic core of Fn3 (Main *et al.*, 1992). It is only four residues away from the mutated residues of the FG loop (Table 1). The results are remarkable because, despite there being 7 mutations and 3 deletions in the loop (more than 10% of total residues; Fig. 12, Table 2), D1.3-4 retains a 3D structure virtually identical to that of the wild-type (except for the mutated loop). Therefore, the results provide strong support that the FG loop is not significantly contributing to the folding and stability of the Fn3 molecule and thus that the FG loop can be mutated extensively.

Table 2. Sequences of oligonucleotides

Name	Sequence
FN1F	CGGGATCCCATATGCAGGTTTCTGATGTTCCGCGTGACC TGGAAGTTGTTGCTGCGACC (SEQ ID NO:13)
FN1R	TAACTGCAGGAGCATCCCAGCTGATCAGCAGGCTAGTC GGGGTCGCAGCAACAAC (SEQ ID NO:14)
FN2F	CTCCTGCAGTTACCGTGCGTTATTACCGTATCACGTACG GTGAAACCGGTG (SEQ ID NO:15)
FN2R	GTGAATTCCTGAACCGGGGAGTTACCACCGGTTTCACC G (SEQ ID NO:16)
FN3F	AGGAATTCACTGTACCTGGTTCCAAGTCTACTGCTACCA TCAGCGG (SEQ ID NO:17)
FN3R	GTATAGTCGACACCCGGTTTCAGGCCGCTGATGGTAGC

(SEQ ID NO:18)

FN4F CGGGTGTCGACTATACCATCACTGTATACGCT (SEQ ID NO:19)

FN4R CGGGATCCGAGCTCGCTGGGCTGTCACCACGGCCAGTA ACAGCGTATACAGTGAT (SEQ ID NO:20)

FN5F CAGCGAGCTCCAAGCCAATCTCGATTAACTACCGT (SEQ ID NO:21)

FN5R CGGGATCCTCGAGTTACTAGGTACGGTAGTTAATCGA (SEQ ID NO:22)

FN5R' CGGGATCCACGCGTGCCACCGGTACGGTAGTTAATCGA (SEQ ID NO:23)

gene3F CGGGATCCACGCGTCCATTCGTTTGTGAATATCAAGGCC AATCG (SEQ ID NO:24)

gene3R CCGGAAGCTTTAAGACTCCTTATTACGCAGTATGTTAGC (SEQ ID NO:25)

38TAABglII CTGTTACTGGCCGTGAGATCTAACCAGCGAGCTCCA (SEQ ID NO:26)

BC3 GATCAGCTGGGATGCTCCTNNKNNKNNKNNKNNKTATT ACCGTATCACGTA (SEQ ID NO:27)

FG2 TGTATACGCTGTTACTGGCNNKNNKNNKNNKNNKNNKN NKTCCAAGCCAATCTCGAT (SEQ ID NO:28)

FG3 CTGTATACGCTGTTACTGGCNNKNNKNNKNNKCCAGCG AGCTCCAAG (SEQ ID NO:29)

FG4 CATCACTGTATACGCTGTTACTNNKNNKNNKNNKNNKT CCAAGCCAATCTC (SEQ ID NO:30)

Restriction enzyme sites are underlined. N and K denote an equimolar mixture of A, T. G and C and that of G and T, respectively.

Structure and stability measurements

Structures of Abs were analyzed using quantitative methods (*e.g.*, DSSP (Kabsch & Sander, 1983) and PDBfit (D. McRee, The Scripps Research

Institute)) as well as computer graphics (*e.g.*, Quanta (Molecular Simulations) and What if (G. Vriend, European Molecular Biology Laboratory)) to superimpose the strand-loop-strand structures of Abs and Fn3.

The stability of monobodies was determined by measuring temperature- and chemical denaturant-induced unfolding reactions (Pace *et al.*, 1989). The temperature-induced unfolding reaction was measured using a circular dichroism (CD) polarimeter. Ellipticity at 222 and 215 nm was recorded as the sample temperature was slowly raised. Sample concentrations between 10 and 50 μM were used. After the unfolding baseline was established, the temperature was lowered to examine the reversibility of the unfolding reaction. Free energy of unfolding was determined by fitting data to the equation for the two-state transition (Becktel & Schellman, 1987; Pace *et al.*, 1989). Nonlinear least-squares fitting was performed using the program Igor (WaveMetrics) on a Macintosh computer.

The structure and stability of two selected mutant Fn3s were studied; the first mutant was D1.3-4 (Table 2) and the second was a mutant called AS40 which contains four mutations in the BC loop ($A^{26}V^{27}T^{28}V^{29}$) → TQRQ). AS40 was randomly chosen from the BC loop library described above. Both mutants were expressed as soluble proteins in *E. coli* and were concentrated at least to 1 mM, permitting NMR studies.

The mid-point of the thermal denaturation for both mutants was approximately 69°C, as compared to approximately 79°C for the wild-type protein. The results indicated that the extensive mutations at the two surface loops did not drastically decrease the stability of Fn3, and thus demonstrated the feasibility of introducing a large number of mutations in both loops.

Stability was also determined by guanidinium chloride (GdnCl)- and urea-induced unfolding reactions. Preliminary unfolding curves were recorded using a fluorometer equipped with a motor-driven syringe; GdnCl or urea were added continuously to the protein solution in the cuvette. Based on the preliminary unfolding curves, separate samples containing varying concentration of a denaturant were prepared and fluorescence (excitation at 290 nm, emission at 300-400 nm) or CD (ellipticity at 222 and 215 nm) were measured after the

samples were equilibrated at the measurement temperature for at least one hour. The curve was fitted by the least-squares method to the equation for the two-state model (Santoro & Bolen, 1988; Koide *et al.*, 1993). The change in protein concentration was compensated if required.

Once the reversibility of the thermal unfolding reaction is established, the unfolding reaction is measured by a Microcal MC-2 differential scanning calorimeter (DSC). The cell (~ 1.3 ml) will be filled with FnAb solution (0.1 - 1 mM) and ΔCp (= $\Delta H/\Delta T$) will be recorded as the temperature is slowly raised. T_m (the midpoint of unfolding), ΔH of unfolding and ΔG of unfolding is determined by fitting the transition curve (Privalov & Potekhin, 1986) with the Origin software provided by Microcal.

Thermal unfolding

A temperature-induced unfolding experiment on Fn3 was performed using circular dichroism (CD) spectroscopy to monitor changes in secondary structure. The CD spectrum of the native Fn3 shows a weak signal near 222 nm (Fig. 3A), consistent with the predominantly β-structure of Fn3 (Perczel *et al.*, 1992). A cooperative unfolding transition is observed at 80-90°C, clearly indicating high stability of Fn3 (Fig. 3B). The free energy of unfolding could not be determined due to the lack of a post-transition baseline. The result is consistent with the high stability of the first Fn3 domain of human fibronectin (Litvinovich *et al.*, 1992), thus indicating that Fn3 domains are in general highly stable.

Binding assays

The binding reactions of monobodies were characterized quantitatively using an isothermal titration calorimeter (ITC) and fluorescence spectroscopy.

The enthalpy change (ΔH) of binding were measured using a Microcal Omega ITC (Wiseman *et al.*, 1989). The sample cell (~ 1.3 ml) was filled with Monobody solution (≤ 100 μM, changed according to K_d), and the reference cell filled with distilled water; the system was equilibrated at a given temperature until a stable baseline is obtained; 5-20 μl of ligand solution (≤ 2 mM) was

injected by a motor-driven syringe within a short duration (20 sec) followed by an equilibration delay (4 minutes); the injection was repeated and heat generation/absorption for each injection was measured. From the change in the observed heat change as a function of ligand concentration, ΔH and K_d was determined (Wiseman *et al.*, 1989). ΔG and ΔS of the binding reaction was deduced from the two directly measured parameters. Deviation from the theoretical curve was examined to assess nonspecific (multiple-site) binding. Experiments were also be performed by placing a ligand in the cell and titrating with an FnAb. It should be emphasized that only ITC gives direct measurement of ΔH, thereby making it possible to evaluate enthalpic and entropic contributions to the binding energy. ITC was successfully used to monitor the binding reaction of the D1.3 Ab (Tello *et al.*, 1993; Bhat *et al.*, 1994).

Intrinsic fluorescence is monitored to measure binding reactions with K_d in the sub-μM range where the determination of K_d by ITC is difficult. Trp fluorescence (excitation at ~ 290 nm, emission at 300-350 nm) and Tyr fluorescence (excitation at ~ 260 nm, emission at ~ 303 nm) is monitored as the Fn3-mutant solution ($\leq$ 10 μM) is titrated with ligand solution ($\leq$ 100 μM). K_d of the reaction is determined by the nonlinear least-squares fitting of the bimolecular binding equation. Presence of secondary binding sites is examined using Scatchard analysis. In all binding assays, control experiments are performed busing wild-type Fn3 (or unrelated monobodies) in place of monobodies of interest.

II. Production of Fn3 mutants with high affinity and specificity Monobodies

Library screening was carried out in order to select monobodies that bind to specific ligands. This is complementary to the modeling approach described above. The advantage of combinatorial screening is that one can easily produce and screen a large number of variants ($\geq 10^8$), which is not feasible with specific mutagenesis ("rational design") approaches. The phage display technique (Smith, 1985; O'Neil & Hoess, 1995) was used to effect the screening processes. Fn3 was fused to a phage coat protein (pIII) and displayed on the surface of

filamentous phages. These phages harbor a single-stranded DNA genome that contains the gene coding the Fn3 fusion protein. The amino acid sequence of defined regions of Fn3 were randomized using a degenerate nucleotide sequence, thereby constructing a library. Phages displaying Fn3 mutants with desired binding capabilities were selected *in vitro*, recovered and amplified. The amino acid sequence of a selected clone can be identified readily by sequencing the Fn3 gene of the selected phage. The protocols of Smith (Smith & Scott, 1993) were followed with minor modifications.

The objective was to produce Monobodies which have high affinity to small protein ligands. HEL and the B1 domain of staphylococcal protein G (hereafter referred to as protein G) were used as ligands. Protein G is small (56 amino acids) and highly stable (Minor & Kim, 1994; Smith *et al.*, 1994). Its structure was determined by NMR spectroscopy (Gronenborn *et al.*, 1991) to be a helix packed against a four-strand β-sheet. The resulting FnAb-protein G complexes (~ 150 residues) is one of the smallest protein-protein complexes produced to date, well within the range of direct NMR methods. The small size, the high stability and solubility of both components and the ability to label each with stable isotopes (^{13}C and ^{15}N; see below for protein G) make the complexes an ideal model system for NMR studies on protein-protein interactions.

The successful loop replacement of Fn3 (the mutant D1.3-4) demonstrate that at least ten residues can be mutated without the loss of the global fold. Based on this, a library was first constructed in which only residues in the FG loop are randomized. After results of loop replacement experiments on the BC loop were obtained, mutation sites were extended that include the BC loop and other sites.

Construction of Fn3 phage display system

An M13 phage-based expression vector pASM1 has been constructed as follows: an oligonucleotide coding the signal peptide of OmpT was cloned at the 5′ end of the Fn3 gene; a gene fragment coding the C-terminal domain of M13 pIII was prepared from the wild-type gene III gene of M13 mp18 using PCR (Corey *et al.*, 1993) and the fragment was inserted at the 3′ end of the

OmpT-Fn3 gene; a spacer sequence has been inserted between Fn3 and pIII. The resultant fragment (OmpT-Fn3-pIII) was cloned in the multiple cloning site of M13 mp18, where the fusion gene is under the control of the lac promoter. This system will produce the Fn3-pIII fusion protein as well as the wild-type pIII protein. The co-expression of wild-type pIII is expected to reduce the number of fusion pIII protein, thereby increasing the phage infectivity (Corey *et al.*, 1993) (five copies of pIII are present on a phage particle). In addition, a smaller number of fusion pIII protein may be advantageous in selecting tight binding proteins, because the chelating effect due to multiple binding sites should be smaller than that with all five copies of fusion pIII (Bass *et al.*, 1990). This system has successfully displayed the serine protease trypsin (Corey *et al.*, 1993). Phages were produced and purified using *E. coli* K91kan (Smith & Scott, 1993) according to a standard method (Sambrook *et al.*, 1989) except that phage particles were purified by a second polyethylene glycol precipitation and acid precipitation.

Successful display of Fn3 on fusion phages has been confirmed by ELISA using an Ab against fibronectin (Sigma), clearly indicating that it is feasible to construct libraries using this system.

An alternative system using the fUSE5 (Parmley & Smith, 1988) may also be used. The Fn3 gene is inserted to fUSE5 using the SfiI restriction sites introduced at the 5′- and 3′- ends of the Fn3 gene PCR. This system displays only the fusion pIII protein (up to five copies) on the surface of a phage. Phages are produced and purified as described (Smith & Scott, 1993). This system has been used to display many proteins and is robust. The advantage of fUSE5 is its low toxicity. This is due to the low copy number of the replication form (RF) in the host, which in turn makes it difficult to prepare a sufficient amount of RF for library construction (Smith & Scott, 1993).

Construction of libraries

The first library was constructed of the Fn3 domain displayed on the surface of M13 phage in which seven residues (77-83) in the FG loop (Fig. 4D) were randomized. Randomization will be achieved by the use of an

oligonucleotide containing degenerated nucleotide sequence. A double-stranded nucleotide was prepared by the same protocol as for gene synthesis (see above) except that one strand had an $(NNK)_6(NNG)$ sequence at the mutation sites, where N corresponds to an equimolar mixture of A, T, G and C and K corresponds to an equimolar mixture of G and T. The (NNG) codon at residue 83 was required to conserve the SacI restriction site (Fig. 2). The (NNK) codon codes all of the 20 amino acids, while the NNG codon codes 14. Therefore, this library contained ~ 10^9 independent sequences. The library was constructed by ligating the double-stranded nucleotide into the wild-type phage vector, pASM1, and the transfecting *E. coli* XL1 blue (Stratagene) using electroporation. XL1 blue has the $lacI^q$ phenotype and thus suppresses the expression of the Fn3-pIII fusion protein in the absence of lac inducers. The initial library was propagated in this way, to avoid selection against toxic Fn3-pIII clones. Phages displaying the randomized Fn3-pIII fusion protein were prepared by propagating phages with K91kan as the host. K91kan does not suppress the production of the fusion protein, because it does not have $lacI^q$. Another library was also generated in which the BC loop (residues 26-20) was randomized.

Selection of displayed Monobodies

Screening of Fn3 phage libraries was performed using the biopanning protocol (Smith & Scott, 1993); a ligand is biotinylated and the strong biotin-streptavidin interaction was used to immobilize the ligand on a streptavidin-coated dish. Experiments were performed at room temperature (~ 22°C). For the initial recovery of phages from a library, 10 μg of a biotinylated ligand were immobilized on a streptavidin-coated polystyrene dish (35 mm, Falcon 1008) and then a phage solution (containing ~ 10^{11} pfu (plaque-forming unit)) was added. After washing the dish with an appropriate buffer (typically TBST, Tris-HCl (50 mM, pH 7.5), NaCl (150 mM) and Tween 20 (0.5%)), bound phages were eluted by one or combinations of the following conditions: low pH, an addition of a free ligand, urea (up to 6 M) and, in the case of anti-protein G Monobodies, cleaving the protein G-biotin linker by thrombin. Recovered phages were amplified using the standard protocol using K91kan as the host

(Sambrook *et al.*, 1989). The selection process were repeated 3-5 times to concentrate positive clones. From the second round on, the amount of the ligand were gradually decreased (to ~ 1 μg) and the biotinylated ligand were mixed with a phage solution before transferring a dish (G. P. Smith, personal communication). After the final round, 10-20 clones were picked, and their DNA sequence will be determined. The ligand affinity of the clones were measured first by the phage-ELISA method (see below).

To suppress potential binding of the Fn3 framework (background binding) to a ligand, wild-type Fn3 may be added as a competitor in the buffers. In addition, unrelated proteins (*e.g.*, bovine serum albumin, cytochrome c and RNase A) may be used as competitors to select highly specific Monobodies.

Binding assay

The binding affinity of Monobodies on phage surface is characterized semi-quantitatively using the phage ELISA technique (Li *et al.*, 1995). Wells of microtiter plates (Nunc) are coated with a ligand protein (or with streptavidin followed by the binding of a biotinylated ligand) and blocked with the Blotto solution (Pierce). Purified phages (~ 10^{10} pfu) originating from single plaques (M13)/colonies (fUSE5) are added to each well and incubated overnight at 4°C. After washing wells with an appropriate buffer (see above), bound phages are detected by the standard ELISA protocol using anti-M13 Ab (rabbit, Sigma) and anti-rabbit Ig-peroxidase conjugate (Pierce) or using anti-M13 Ab-peroxidase conjugate (Pharmacia). Colormetric assays are performed using TMB (3,3′,5,5′-tetramethylbenzidine, Pierce). The high affinity of protein G to immunoglobulins present a special problem; Abs cannot be used in detection. Therefore, to detect anti-protein G Monobodies, fusion phages are immobilized in wells and the binding is then measured using biotinylated protein G followed by the detection using streptavidin-peroxidase conjugate.

Production of soluble Monobodies

After preliminary characterization of mutant Fn3s using phage ELISA, mutant genes are subcloned into the expression vector pEW1. Mutant proteins

are produced as His•tag fusion proteins and purified, and their conformation, stability and ligand affinity are characterized.

III. Increased Stability of Fn3 Scaffolds

The definition of "higher stability" of a protein is the ability of a protein to retain its three-dimensional structure required for function at a higher temperature (in the case of thermal denaturation), and in the presence of a higher concentration of a denaturing chemical reagent such as guanidine hydrochloride. This type of "stability" is generally called "conformational stability." It has been shown that conformational stability is correlated with resistance against proteolytic degradation, *i.e.*, breakdown of protein in the body (Kamtekar *et al.* 1993).

Improving the conformational stability is a major goal in protein engineering. Here, mutations have been developed by the inventor that enhance the stability of the fibronectin type III domain (Fn3). The inventor has developed a technology in which Fn3 is used as a scaffold to engineer artificial binding proteins (Koide *et al.*, 1998). It has been shown that many residues in the surface loop regions of Fn3 can be mutated without disrupting the overall structure of the Fn3 molecule, and that variants of Fn3 with a novel binding function can be engineered using combinatorial library screening (Koide *et al.*, 1998). The inventor found that, although Fn3 is an excellent scaffold, Fn3 variants that contain large number of mutations are destablized against chemical denaturation, compared to the wild-type Fn3 protein (Koide *et al.*, 1998). Thus, as the number of mutated positions are mutated in order to engineer a new binding function, the stability of such Fn3 variants further decreases, ultimately leading to marginally stable proteins. Because artificial binding proteins must maintain their three-dimensional structure to be functional, stability limits the number of mutations that can be introduced in the scaffold. Thus, modifications of the Fn3 scaffold that increase its stability are useful in that they allow one to introduce more mutations for better function, and that they make it possible to use Fn3-based engineered proteins in a wider range of applications.

The inventor found that wild-type Fn3 is more stable at acidic pH than at

neutral pH (Koide *et al.*, 1998). The pH dependence of Fn3 stability is characterized in Figure 18. The pH dependence curve has an apparent transition midpoint near pH 4 (Figure 18). These results suggest that by identifying and removing destablizing interactions in Fn3 one is able to improve the stability of Fn3 at neutral pH. It should be noted that most applications of engineered Fn3, such as diagnostics, therapeutics and catalysts, are expected to be used near neutral pH, and thus it is important to improve the stability at neutral pH. Studies by other investigators have demonstrated that the optimization of surface electrostatic properties can lead to a substantial increase in protein stability (Perl *et al.* 2000, Spector *et al.* 1999, Loladze *et al.* 1999, Grimsley *et al.* 1999).

The pH dependence of Fn3 stability suggests that amino acids with pK_a near 4 are involved in the observed transition. The carboxyl groups of aspartic acid (Asp) and glutamic acid (Glu) have pK_a in this range (Creighton, T.E. 1993). It is well known that if a carboxyl group has unfavorable (*i.e.* destabilizing) interactions in a protein, its pK_a is shifted to a higher value from its standard, unperturbed value (Yang and Honig 1992). Thus, the pK_a values of all carboxyl groups in Fn3 were determined using nuclear magnetic resonance (NMR) spectrosocpy, to identify carboxyl groups with unusual pK_a's, as shown below.

First, the ^{13}C resonance for the carboxyl carbon of each Asp and Glu residue were assigned (Figure 19). Next pH titration of ^{13}C resonances was performed for these groups (Figure 20). The pK_a values for these residues are listed in Table 3.

Table 3. pK_a values for Asp and Glu residues in Fn3.

Residue	pK_a
E9	5.09
E38	3.79
E47	3.94
D3	3.66
D7	3.54, 5.54*
D23	3.54, 5.25*

D67	4.18
D80	3.40

The standard deviation in the pK_a values are less than 0.05 pH units.

*Data for D7 and D23 were fitted with a transition curve with two pK_a values.

These results show that Asp 7 and 23, and Glu 9 have up-shifted pK_a's with respect to their unperturbed pK_a's (approximately 4.0), indicating that these residues are involved in unfavorable interactions. In contrast, the other Asp and Glu residues have pK_a's close to the respective unperturbed values, indicating that the carboxyl groups of these residues do not significantly contribute to the stability of Fn3.

In the three-dimensional structure of Fn3 (Main *et al.* 1992), Asp 7 and 23, and Glu 9 form a patch on the surface (Figure 21), with Asp 7 centrally located in the patch. This spatial proximity of these negatively charged residues explains why these residues have unfavorable interactions in Fn3. At low pH where these residues are protonated and neutral, the unfavorable interactions are expected to be mostly relieved. At the same time, the structure suggests that the stability of Fn3 at neutral pH could be improved if the electrostatic repulsion between these three residues is removed. Because Asp 7 is centrally located among the three residues, it was decided to mutate Asp 7. Two mutants were prepared, D7N and D7K (*i.e.*, the aspartic acid at amino acid residue number 7 was substituted with an asparagine residue or a lysine residue, respectively). The former replaces the negative charge with a neutral residue of virtually the same size. The latter places a positive charge at residue 7.

The degrees of stability of the mutant proteins were characterized in thermal and chemical denaturation measurements. In thermal denaturation measurements, denaturation of the Fn3 proteins was monitored using circular dichroism spectroscopy at the wavelength of 227 nm. All the proteins underwent a cooperative transition (Figure 22). From the transition curves, the midpoints of the transition (T_m) for the wild-type, D7N and D7K were determined to be 62, 69 and 70 °C in 0.02 M sodium phosphate buffer (pH 7.0) containing 0.1 M sodium chloride and 6.2 M urea. Thus, the mutations

increased the T_m of wild-type Fn3 by 7-8 °C.

Chemical denaturation of Fn3 proteins was monitored using fluorescence emission from the single Trp residue of Fn3 (Figure 23). The free energies of unfolding in the absence of guanidine HCl (ΔG^0) were determined to be 7.4, 8.1 and 8.0 kcal/mol for the wild-type, D7N and D7K, respectively (a larger ΔG^0 indicates a higher stability). The two mutants were again found to be more stable than the wild-type protein.

These results show that a point mutation on the surface can significantly enhance the stability of Fn3. Because these mutations are on the surface, they minimally alter the structure of Fn3, and they can be easily introduced to other, engineered Fn3 proteins. In addition, mutations at Glu 9 and/or Asp 23 also enhance the stability of Fn3. Furthermore, mutations at one or more of these three residues can be combined.

Thus, Fn3 is the fourth example of a monomeric immunoglobulin-like scaffold that can be used for engineering binding proteins. Successful selection of novel binding proteins have also been based on minibody, tendamistat and "camelized" immunoglobulin VH domain scaffolds (Martin *et al.*, 1994; Davies & Riechmann, 1995; McConnell & Hoess, 1995). The Fn3 scaffold has advantages over these systems. Bianchi *et al.* reported that the stability of a minibody was 2.5 kcal/mol, significantly lower than that of Ubi4-K. No detailed structural characterization of minibodies has been reported to date. Tendamistat and the VH domain contain disulfide bonds, and thus preparation of correctly folded proteins may be difficult. Davies and Riechmann reported that the yields of their camelized VH domains were less than 1 mg per liter culture (Davies & Riechmann, 1996).

Thus, the Fn3 framework can be used as a scaffold for molecular recognition. Its small size, stability and well-characterized structure make Fn3 an attractive system. In light of the ubiquitous presence of Fn3 in a wide variety of natural proteins involved in ligand binding, one can engineer Fn3-based binding proteins to different classes of targets.

The following examples are intended to illustrate but not limit the invention.

EXAMPLE I

Construction of the Fn3 gene

A synthetic gene for tenth Fn3 of fibronectin (Fig.1) was designed on the basis of amino acid residue 1416-1509 of human fibronectin (Kornblihtt, et al., 1985) and its three dimensional structure (Main, *et al.*, 1992). The gene was engineered to include convenient restriction sites for mutagenesis and the so-called "preferred codons" for high level protein expression (Gribskov, *et al.*, 1984) were used. In addition, a glutamine residue was inserted after the N-terminal methionine in order to avoid partial processing of the N-terminal methionine which often degrades NMR spectra (Smith, *et al.*, 1994). Chemical reagents were of the analytical grade or better and purchased from Sigma Chemical Company and J.T. Baker, unless otherwise noted. Recombinant DNA procedures were performed as described in "Molecular Cloning" (Sambrook, *et al.*, 1989), unless otherwise stated. Custom oligonucleotides were purchased from Operon Technologies. Restriction and modification enzymes were from New England Biolabs.

The gene was assembled in the following manner. First, the gene sequence (Fig. 5) was divided into five parts with boundaries at designed restriction sites: fragment 1, NdeI-PstI (oligonucleotides FN1F and FN1R (Table 2); fragment 2, PstI-EcoRI (FN2F and FN2R); fragment 3, EcoRI-SalI (FN3F and FN3R); fragment 4, SalI-SacI (FN4F and FN4R); fragment 5, SacI-BamHI (FN5F and FN5R). Second, for each part, a pair of oligonucleotides which code opposite strands and have complementary overlaps of approximately 15 bases was synthesized. These oligonucleotides were designated FN1F-FN5R and are shown in Table 2. Third, each pair (*e.g.*, FN1F and FN1R) was annealed and single-strand regions were filled in using the Klenow fragment of DNA polymerase. Fourth, the double stranded oligonucleotide was digested with the relevant restriction enzymes at the termini of the fragment and cloned into the pBlueScript SK plasmid (Stratagene) which had been digested with the same enzymes as those used for the fragments. The DNA sequence of the inserted fragment was confirmed by DNA sequencing using an Applied Biosystems DNA sequencer and the dideoxy termination protocol provided by the manufacturer.

Last, steps 2-4 were repeated to obtain the entire gene.

The gene was also cloned into the pET3a and pET15b (Novagen) vectors (pAS45 and pAS25, respectively). The maps of the plasmids are shown in Figs. 6 and 7. *E. coli* BL21 (DE3) (Novagen) containing these vectors expressed the Fn3 gene under the control of bacteriophage T7 promotor (Studier, *et al.*, 1990); pAS24 expresses the 96-residue Fn3 protein only, while pAS45 expresses Fn3 as a fusion protein with poly-histidine peptide (His•tag). High level expression of the Fn3 protein and its derivatives in *E. coli* was detected as an intense band on SDS-PAGE stained with CBB.

The binding reaction of the monobodies is characterized quantitatively by means of fluorescence spectroscopy using purified soluble monobodies.

Intrinsic fluorescence is monitored to measure binding reactions. Trp fluorescence (excitation at ~290 nm, emission at 300 350 nm) and Tyr fluorescence (excitation at ~260 nm, emission at ~303 nm) is monitored as the Fn3-mutant solution (≤ 100 μM) is titrated with a ligand solution. When a ligand is fluorescent (*e.g.* fluorescein), fluorescence from the ligand may be used. K_d of the reaction will be determined by the nonlinear least-squares fitting of the bimolecular binding equation.

If intrinsic fluorescence cannot be used to monitor the binding reaction, monobodies are labeled with fluorescein-NHS (Pierce) and fluorescence polarization is used to monitor the binding reaction (Burke *et al.*, 1996).

EXAMPLE II

Modifications to include restriction sites in the Fn3 gene

The restriction sites were incorporated in the synthetic Fn3 gene without changing the amino acid sequence Fn3. The positions of the restriction sites were chosen so that the gene construction could be completed without synthesizing long (>60 bases) oligonucleotides and so that two loop regions could be mutated (including by randomization) by the cassette mutagenesis method (*i.e.*, swapping a fragment with another synthetic fragment containing mutations). In addition, the restriction sites were chosen so that most sites were unique in the vector for phage display. Unique restriction sites allow one to

recombine monobody clones which have been already selected in order to supply a larger sequence space.

EXAMPLE III

Construction of M13 phage display libraries

A vector for phage display, pAS38 (for its map, see Fig. 8) was constructed as follows. The XbaI-BamHI fragment of pET12a encoding the signal peptide of OmpT was cloned at the 5' end of the Fn3 gene. The C-terminal region (from the FN5F and FN5R oligonucleotides, see Table 2) of the Fn3 gene was replaced with a new fragment consisting of the FN5F and FN5R' oligonucleotides (Table 2) which introduced a MluI site and a linker sequence for making a fusion protein with the pIII protein of bacteriophage M13. A gene fragment coding the C-terminal domain of M13 pIII was prepared from the wild-type gene III of M13mp18 using PCR (Corey, *et al.*, 1993) and the fragment was inserted at the 3' end of the OmpT-Fn3 fusion gene using the MluI and HindIII sites.

Phages were produced and purified using a helper phage, M13K07, according to a standard method (Sambrook, *et al.*, 1989) except that phage particles were purified by a second polyethylene glycol precipitation. Successful display of Fn3 on fusion phages was confirmed by ELISA (Harlow & Lane, 1988) using an antibody against fibronectin (Sigma) and a custom anti-FN3 antibody (Cocalico Biologicals, PA, USA).

EXAMPLE IV

Libraries containing loop variegations in the AB loop

A nucleic acid phage display library having variegation in the AB loop is prepared by the following methods. Randomization is achieved by the use of oligonucleotides containing degenerated nucleotide sequence. Residues to be variegated are identified by examining the X-ray and NMR structures of Fn3 (Protein Data Bank accession numbers, 1FNA and 1TTF, respectively). Oligonucleotides containing NNK (N and K here denote an equimolar mixture of A, T, G, and C and an equimolar mixture of G and T, respectively) for the

variegated residues are synthesized (see oligonucleotides BC3, FG2, FG3, and FG4 in Table 2 for example). The NNK mixture codes for all twenty amino acids and one termination codon (TAG). TAG, however, is suppressed in the *E. coli* XL-1 blue. Single-stranded DNAs of pAS38 (and its derivatives) are prepared using a standard protocol (Sambrook, *et al.*, 1989).

Site-directed mutagenesis is performed following published methods (see for example, Kunkel, 1985) using a Muta-Gene kit (BioRad). The libraries are constructed by electroporation of *E. coli* XL-1 Blue electroporation competent cells (200 µl; Stratagene) with 1µg of the plasmid DNA using a BTX electrocell manipulator ECM 395 1mm gap cuvette. A portion of the transformed cells is plated on an LB-agar plate containing ampicillin (100 µg/ml) to determine the transformation efficiency. Typically, 3 X 10^8 transformants are obtained with 1 µg of DNA, and thus a library contains 10^8 to 10^9 independent clones. Phagemid particles were prepared as described above.

EXAMPLE V

Loop variegations in the BC, CD, DE, EF or FG loop

A nucleic acid phage display library having five variegated residues (residues number 26-30) in the BC loop, and one having seven variegated residues (residue numbers 78-84) in the FG loop, was prepared using the methods described in Example IV above. Other nucleic acid phage display libraries having variegation in the CD, DE or EF loop can be prepared by similar methods.

EXAMPLE VI

Loop variegations in the FG and BC loop

A nucleic acid phage display library having seven variegated residues (residues number 78-84) in the FG loop and five variegated residues (residue number 26-30) in the BC loop was prepared. Variegations in the BC loop were prepared by site-directed mutagenesis (Kunkel, *et al.*) using the BC3 oligonucleotide described in Table 1. Variegations in the FG loop were introduced using site-directed mutagenesis using the BC loop library as the

starting material, thereby resulting in libraries containing variegations in both BC and FG loops. The oligonucleotide FG2 has variegating residues 78-84 and oligonucleotide FG4 has variegating residues 77-81 and a deletion of residues 82-84.

A nucleic acid phage display library having five variegated residues (residues 78-84) in the FG loop and a three residue deletion (residues 82-84) in the FG loop, and five variegated residues (residues 26-30) in the BC loop, was prepared. The shorter FG loop was made in an attempt to reduce the flexibility of the FG loop; the loop was shown to be highly flexible in Fn3 by the NMR studies of Main, *et al.* (1992). A highly flexible loop may be disadvantageous to forming a binding site with a high affinity (a large entropy loss is expected upon the ligand binding, because the flexible loop should become more rigid). In addition, other Fn3 domains (besides human) have shorter FG loops (for sequence alignment, see Figure 12 in Dickinson, *et al.* (1994)).

Randomization was achieved by the use of oligonucleotides containing degenerate nucleotide sequence (oligonucleotide BC3 for variegating the BC loop and oligonucleotides FG2 and FG4 for variegating the FG loops).

Site-directed mutagenesis was performed following published methods (see for example, Kunkel, 1985). The libraries were constructed by electrotransforming *E. coli* XL-1 Blue (Stratagene). Typically a library contains 10^8 to 10^9 independent clones. Library 2 contains five variegated residues in the BC loop and seven variegated residues in the FG loop. Library 4 contains five variegated residues in each of the BC and FG loops, and the length of the FG loop was shortened by three residues.

EXAMPLE VII

fd phage display libraries constructed with loop variegations

Phage display libraries are constructed using the fd phage as the genetic vector. The Fn3 gene is inserted in fUSE5 (Parmley & Smith, 1988) using SfiI restriction sites which are introduced at the 5′ and 3′ ends of the Fn3 gene using PCR. The expression of this phage results in the display of the fusion pIII protein on the surface of the fd phage. Variegations in the Fn3 loops are

introduced using site-directed mutagenesis as described hereinabove, or by subcloning the Fn3 libraries constructed in M13 phage into the fUSE5 vector.

EXAMPLE VIII

Other phage display libraries

T7 phage libraries (Novagen, Madison, WI) and bacterial pili expression systems (Invitrogen) are also useful to express the Fn3 gene.

EXAMPLE IX

Isolation of polypeptides which bind to macromolecular structures

The selection of phage-displayed monobodies was performed following the protocols of Barbas and coworkers (Rosenblum & Barbas, 1995). Briefly, approximately 1 μg of a target molecule ("antigen") in sodium carbonate buffer (100 mM, pH 8.5) was immobilized in the wells of a microtiter plate (Maxisorp, Nunc) by incubating overnight at 4°C in an air tight container. After the removal of this solution, the wells were then blocked with a 3% solution of BSA (Sigma, Fraction V) in TBS by incubating the plate at 37°C for 1 hour. A phagemid library solution (50 μl) containing approximately 10^{12} colony forming units (cfu) of phagemid was absorbed in each well at 37°C for 1 hour. The wells were then washed with an appropriate buffer (typically TBST, 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 0.5% Tween20) three times (once for the first round). Bound phage were eluted by an acidic solution (typically, 0.1 M glycine-HCl, pH 2.2; 50 μl) and recovered phage were immediately neutralized with 3 μl of Tris solution. Alternatively, bound phage were eluted by incubating the wells with 50 μl of TBS containing the antigen (1 - 10 μM). Recovered phage were amplified using the standard protocol employing the XL1Blue cells as the host (Sambrook, *et al.*). The selection process was repeated 5-6 times to concentrate positive clones. After the final round, individual clones were picked and their binding affinities and DNA sequences were determined.

The binding affinities of monobodies on the phage surface were characterized using the phage ELISA technique (Li, *et al.*, 1995). Wells of microtiter plates (Nunc) were coated with an antigen and blocked with BSA.

Purified phages (10^8- 10^{11} cfu) originating from a single colony were added to each well and incubated 2 hours at 37°C. After washing wells with an appropriate buffer (see above), bound phage were detected by the standard ELISA protocol using anti-M13 antibody (rabbit, Sigma) and anti-rabbit Ig-peroxidase conjugate (Pierce). Colorimetric assays were performed using Turbo-TMB (3,3',5,5'-tetramethylbenzidine, Pierce) as a substrate.

The binding affinities of monobodies on the phage surface were further characterized using the competition ELISA method (Djavadi-Ohaniance, *et al.*, 1996). In this experiment, phage ELISA is performed in the same manner as described above, except that the phage solution contains a ligand at varied concentrations. The phage solution was incubated a 4°C for one hour prior to the binding of an immobilized ligand in a microtiter plate well. The affinities of phage displayed monobodies are estimated by the decrease in ELISA signal as the free ligand concentration is increased.

After preliminary characterization of monobodies displayed on the surface of phage using phage ELISA, genes for positive clones were subcloned into the expression vector pAS45. *E. coli* BL21(DE3) (Novagen) was transformed with an expression vector (pAS45 and its derivatives). Cells were grown in M9 minimal medium and M9 medium supplemented with Bactotryptone (Difco) containing ampicillin (200 μg/ml). For isotopic labeling, ^{15}N NH_4Cl and/or ^{13}C glucose replaced unlabeled components. Stable isotopes were purchased from Isotec and Cambridge Isotope Labs. 500 ml medium in a 2 l baffle flask was inoculated with 10 ml of overnight culture and agitated at approximately 140 rpm at 37°C. IPTG was added at a final concentration of 1 mM to induce protein expression when OD(600 nm) reached approximately 1.0. The cells were harvested by centrifugation 3 hours after the addition of IPTG and kept frozen at -70°C until used.

Fn3 and monobodies with His•tag were purified as follows. Cells were suspended in 5 ml/(g cell) of 50 mM Tris (pH 7.6) containing 1 mM phenylmethylsulfonyl fluoride. HEL (Sigma, 3X crystallized) was added to a final concentration of 0.5 mg/ml. After incubating the solution for 30 min at 37°C, it was sonicated so as to cause cell breakage three times for 30 seconds on

ice. Cell debris was removed by centrifugation at 15,000 rpm in an Sorval RC-2B centrifuge using an SS-34 rotor. Concentrated sodium chloride is added to the solution to a final concentration of 0.5 M. The solution was then applied to a 1 ml HisTrap™ chelating column (Pharmacia) preloaded with nickel chloride (0.1 M, 1 ml) and equilibrated in the Tris buffer (50 mM, pH 8.0) containing 0.5 M sodium chloride. After washing the column with the buffer, the bound protein was eluted with a Tris buffer (50 mM, pH 8.0) containing 0.5 M imidazole. The His•tag portion was cleaved off, when required, by treating the fusion protein with thrombin using the protocol supplied by Novagen (Madison, WI). Fn3 was separated from the His•tag peptide and thrombin by a Resources®column (Pharmacia) using a linear gradient of sodium chloride (0 - 0.5 M) in sodium acetate buffer (20 mM, pH 5.0).

Small amounts of soluble monobodies were prepared as follows. XL-1 Blue cells containing pAS38 derivatives (plasmids coding Fn3-pIII fusion proteins) were grown in LB media at 37°C with vigorous shaking until OD(600 nm) reached approximately 1.0; IPTG was added to the culture to a final concentration of 1 mM, and the cells were further grown overnight at 37°C. Cells were removed from the medium by centrifugation, and the supernatant was applied to a microtiter well coated with a ligand. Although XL-1 Blue cells containing pAS38 and its derivatives express FN3-pIII fusion proteins, soluble proteins are also produced due to the cleavage of the linker between the Fn3 and pIII regions by proteolytic activities of *E. coli* (Rosenblum & Barbas, 1995). Binding of a monobody to the ligand was examined by the standard ELISA protocol using a custom antibody against Fn3 (purchased from Cocalico Biologicals, Reamstown, PA). Soluble monobodies obtained from the periplasmic fraction of *E. coli* cells using a standard osmotic shock method were also used.

EXAMPLE X

Ubiquitin binding monobody

Ubiquitin is a small (76 residue) protein involved in the degradation pathway in eurkaryotes. It is a single domain globular protein. Yeast ubiquitin was purchased from Sigma Chemical Company and was used without further purification.

Libraries 2 and 4, described in Example VI above, were used to select ubiquitin-binding monobodies. Ubiquitin (1 μg in 50 μl sodium bicarbonate buffer (100 mM, pH 8.5)) was immobilized in the wells of a microtiter plate, followed by blocking with BSA (3% in TBS). Panning was performed as described above. In the first two rounds, 1 μg of ubiquitin was immobilized per well, and bound phage were elute with an acidic solution. From the third to the sixth rounds, 0.1 μg of ubiquitin was immobilized per well and the phage were eluted either with an acidic solution or with TBS containing 10 μM ubiquitin.

Binding of selected clones was tested first in the polyclonal mode, *i.e.*, before isolating individual clones. Selected clones from all libraries showed significant binding to ubiquitin. These results are shown in Figure 9. The binding to the immobilized ubiquitin of the clones was inhibited almost completely by less than 30 μM soluble ubiquitin in the competition ELISA experiments (see Fig. 10). The sequences of the BC and FG loops of ubiquitin-binding monobodies is shown in Table 4.

Table 4. Sequences of ubiquitin-binding monobodies

Name	BC loop	FG loop	Occurrence (if more than one)
211	CARRA (SEQ ID NO:31)	RWIPLAK (SEQ ID NO:32)	2
212	CWRRA (SEQ ID NO:33)	RWVGLAW (SEQ ID NO:34)	
213	CKHRR (SEQ ID NO:35)	FADLWWR (SEQ ID NO:36)	
214	CRRGR (SEQ ID NO:37)	RGFMWLS (SEQ ID NO:38)	
215	CNWRR (SEQ ID NO:39)	RAYRYRW (SEQ ID NO:40)	
411	SRLRR (SEQ ID NO:41)	PPWRV (SEQ ID NO:42)	9
422	ARWTL (SEQ ID NO:43)	RRWWW (SEQ ID NO:44)	
424	GQRTF (SEQ ID NO:45)	RRWWA (SEQ ID NO:46)	

The 411 clone, which was the most enriched clone, was characterized using phage ELISA. The 411 clone showed selective binding and inhibition of binding in the presence of about 10 μM ubiquitin in solution (Fig. 11).

EXAMPLE XI

Methods for the immobilization of small molecules

Target molecules were immobilized in wells of a microtiter plate (Maxisorp, Nunc) as described hereinbelow, and the wells were blocked with BSA. In addition to the use of carrier protein as described below, a conjugate of a target molecule in biotin can be made. The biotinylated ligand can then be immobilized to a microtiter plate well which has been coated with streptavidin.

In addition to the use of a carrier protein as described below, one could make a conjugate of a target molecule and biotin (Pierce) and immobilize a biotinylated ligand to a microtiter plate well which has been coated with streptavidin (Smith and Scott, 1993).

Small molecules may be conjugated with a carrier protein such as bovine

serum albumin (BSA, Sigma), and passively adsorbed to the microtiter plate well. Alternatively, methods of chemical conjugation can also be used. In addition, solid supports other than microtiter plates can readily be employed.

EXAMPLE XII

Fluorescein binding monobody

Fluorescein has been used as a target for the selection of antibodies from combinatorial libraries (Barbas, *et al.* 1992). NHS-fluorescein was obtained from Pierce and used according to the manufacturer's instructions in preparing conjugates with BSA (Sigma). Two types of fluorescein-BSA conjugates were prepared with approximate molar ratios of 17 (fluorescein) to one (BSA).

The selection process was repeated 5-6 times to concentrate positive clones. In this experiment, the phage library was incubated with a protein mixture (BSA, cytochrome C (Sigma, Horse) and RNaseA (Sigma, Bovine), 1 mg/ml each) at room temperature for 30 minutes, prior to the addition to ligand coated wells. Bound phage were eluted in TBS containing 10 μM soluble fluorescein, instead of acid elution. After the final round, individual clones were picked and their binding affinities (see below) and DNA sequences were determined.

Table 5. Clones from Library #2

	BC	FG
WT	AVTVR (SEQ ID NO:47)	RGDSPAS (SEQ ID NO:48)
pLB24.1	CNWRR (SEQ ID NO:49)	RAYRYRW (SEQ ID NO:50)
pLB24.2	CMWRA (SEQ ID NO:51)	RWGMLRR (SEQ ID NO:52)
pLB24.3	ARMRE (SEQ ID NO:53)	RWLRGRY (SEQ ID NO:54)
pLB24.4	CARRR (SEQ ID NO:55)	RRAGWGW (SEQ ID NO:56)
pLB24.5	CNWRR (SEQ ID NO:57)	RAYRYRW (SEQ ID NO:58)
pLB24.6	RWRER (SEQ ID NO:59)	RHPWTER (SEQ ID NO:60)
pLB24.7	CNWRR (SEQ ID NO:61)	RAYRYRW (SEQ ID NO:62)
pLB24.8	ERRVP (SEQ ID NO:63)	RLLLWQR (SEQ ID NO:64)
pLB24.9	GRGAG (SEQ ID NO:65)	FGSFERR (SEQ ID NO:66)
pLB24.11	CRWTR (SEQ ID NO:67)	RRWFDGA (SEQ ID NO:68)
pLB24.12	CNWRR (SEQ ID NO:69)	RAYRYRW (SEQ ID NO:70)

Clones from Library #4

WT	AVTVR (SEQ ID NO:71)	GRGDS (SEQ ID NO:72)
pLB25.1	GQRTF (SEQ ID NO:73)	RRWWA (SEQ ID NO:74)
pLB25.2	GQRTF (SEQ ID NO:75)	RRWWA (SEQ ID NO:76)
pLB25.3	GQRTF (SEQ ID NO:77)	RRWWA (SEQ ID NO:78)
pLB25.4	LRYRS (SEQ ID NO:79)	GWRWR (SEQ ID NO:80)
pLB25.5	GQRTF (SEQ ID NO:81)	RRWWA (SEQ ID NO:82)
pLB25.6	GQRTF (SEQ ID NO:83)	RRWWA (SEQ ID NO:84)
pLB25.7	LRYRS (SEQ ID NO:85)	GWRWR (SEQ ID NO:86)
pLB25.9	LRYRS (SEQ ID NO:87)	GWRWR (SEQ ID NO:88)
pLB25.11	GQRTF (SEQ ID NO:89)	RRWWA (SEQ ID NO:90)
pLB25.12	LRYRS (SEQ ID NO:91)	GWRWR (SEQ ID NO:92)

Preliminary characterization of the binding affinities of selected clones were performed using phage ELISA and competition phage ELISA (see Fig. 12 (Fluorescein-1) and Fig. 13 (Fluorescein-2)). The four clones tested showed

specific binding to the ligand-coated wells, and the binding reactions are inhibited by soluble fluorescein (see Fig. 13).

EXAMPLE XIII

Digoxigenin binding monobody

Digoxigenin-3-O-methyl-carbonyl-e-aminocapronic acid-NHS (Boehringer Mannheim) is used to prepare a digoxigenin-BSA conjugate. The coupling reaction is performed following the manufacturers' instructions. The digoxigenin-BSA conjugate is immobilized in the wells of a microtiter plate and used for panning. Panning is repeated 5 to 6 times to enrich binding clones. Because digoxigenin is sparingly soluble in aqueous solution, bound phages are eluted from the well using acidic solution. See Example XIV.

EXAMPLE XIV

TSAC (transition state analog compound) binding monobodies

Carbonate hydrolyzing monobodies are selected as follows. A transition state analog for carbonate hydrolysis, 4-nitrophenyl phosphonate is synthesized by an Arbuzov reaction as described previously (Jacobs and Schultz, 1987). The phosphonate is then coupled to the carrier protein, BSA, using carbodiimide, followed by exhaustive dialysis (Jacobs and Schultz, 1987). The hapten-BSA conjugate is immobilized in the wells of a microtiter plate and monobody selection is performed as described above. Catalytic activities of selected monobodies are tested using 4-nitrophenyl carbonate as the substrate.

Other haptens useful to produce catalytic monobodies are summarized in H. Suzuki (1994) and in N. R. Thomas (1994).

EXAMPLE XV

NMR characterization of Fn3 and comparison of the Fn3 secreted by yeast with that secreted by *E. coli*

Nuclear magnetic resonance (NMR) experiments are performed to identify the contact surface between FnAb and a target molecule, *e.g.*, monobodies to fluorescein, ubiquitin, RNaseA and soluble derivatives of

digoxigenin. The information is then be used to improve the affinity and specificity of the monobody. Purified monobody samples are dissolved in an appropriate buffer for NMR spectroscopy using Amicon ultrafiltration cell with a YM-3 membrane. Buffers are made with 90 % H_2O/10 % D_2O (distilled grade, Isotec) or with 100 % D_2O. Deuterated compounds (*e.g.* acetate) are used to eliminate strong signals from them.

NMR experiments are performed on a Varian Unity INOVA 600 spectrometer equipped with four RF channels and a triple resonance probe with pulsed field gradient capability. NMR spectra are analyzed using processing programs such as Felix (Molecular Simulations), nmrPipe, PIPP, and CAPP (Garrett, *et al.*, 1991; Delaglio, *et al.*, 1995) on UNIX workstations. Sequence specific resonance assignments are made using well-established strategy using a set of triple resonance experiments (CBCA(CO)NH and HNCACB) (Grzesiek & Bax, 1992; Wittenkind & Mueller, 1993).

Nuclear Overhauser effect (NOE) is observed between ^1H nuclei closer than approximately 5 Å, which allows one to obtain information on interproton distances. A series of double- and triple-resonance experiments (Table 6; for recent reviews on these techniques, see Bax & Grzesiek, 1993 and Kay, 1995) are performed to collect distance (*i.e.* NOE) and dihedral angle (J-coupling) constraints. Isotope-filtered experiments are performed to determine resonance assignments of the bound ligand and to obtain distance constraints within the ligand and those between FnAb and the ligand. Details of sequence specific resonance assignments and NOE peak assignments have been described in detail elsewhere (Clore & Gronenborn, 1991; Pascal, *et al.*, 1994b; Metzler, *et al.*, 1996).

Table 6. NMR experiments for structure characterization

Experiment Name	Reference
1. reference spectra	
2D-^1H, ^{15}N-HSQC	(Bodenhausen & Ruben, 1980; Kay, *et al.*, 1992)

2D-^{1}H, ^{13}C-HSQC (Bodenhausen & Ruben, 1980; Vuister & Bax, 1992)

2. backbone and side chain resonance assignments of ^{13}C/^{15}N-labeled protein

3D-CBCA(CO)NH	(Grzesiek & Bax, 1992)
3D-HNCACB	(Wittenkind & Mueller, 1993)
3D-C(CO)NH	(Logan *et al.*, 1992; Grzesiek *et al.*, 1993)
3D-H(CCO)NH	
3D-HBHA(CBCACO)NH	(Grzesiek & Bax, 1993)
3D-HCCH-TOCSY	(Kay *et al.*, 1993)
3D-HCCH-COSY	(Ikura *et al.*, 1991)
3D-^{1}H, ^{15}N-TOCSY-HSQC	(Zhang *et al.*, 1994)
2D-HB(CBCDCE)HE	(Yamazaki *et al.*, 1993)

3. resonance assignments of unlabeled ligand

2D-isotope-filtered ^{1}H-TOCSY	
2D-isotope-filtered ^{1}H-COSY	
2D-isotope-filtered ^{1}H-NOESY	(Ikura & Bax, 1992)

4. structural constraints

within labeled protein

3D-^{1}H, ^{15}N-NOESY-HSQC	(Zhang *et al.*, 1994)
4D-^{1}H, ^{13}C-HMQC-NOESY-HMQC	(Vuister *et al.*, 1993)
4D-^{1}H, ^{13}C, ^{15}N-HSQC-NOESY-HSQC	(Muhandiram *et al.*, 1993; Pascal *et al.*, 1994a)

within unlabeled ligand

2D-isotope-filtered ^{1}H-NOESY	(Ikura & Bax, 1992)

interactions between protein and ligand

3D-isotope-filtered ^{1}H, ^{15}N-NOESY-HSQC	
3D-isotope-filtered ^{1}H, ^{13}C-NOESY-HSQC	(Lee *et al.*, 1994)

5. dihedral angle constraints

J-molulated ^{1}H, ^{15}N-HSQC	(Billeter *et al.*, 1992)
3D-HNHB	(Archer *et al.*, 1991)

Backbone ^{1}H, ^{15}N and ^{13}C resonance assignments for a monobody are compared to those for wild-type Fn3 to assess structural changes in the mutant. Once these data establish that the mutant retains the global structure, structural refinement is performed using experimental NOE data. Because the structural difference of a monobody is expected to be minor, the wild-type structure can be used as the initial model after modifying the amino acid sequence. The mutations are introduced to the wild-type structure by interactive molecular modeling, and then the structure is energy-minimized using a molecular modeling program such as Quanta (Molecular Simulations). Solution structure is refined using cycles of dynamical simulated annealing (Nilges *et al.*, 1988) in the program X-PLOR (Brünger, 1992). Typically, an ensemble of fifty structures is calculated. The validity of the refined structures is confirmed by calculating a fewer number of structures from randomly generated initial structures in X-PLOR using the YASAP protocol (Nilges, *et al.*, 1991). Structure of a monobody-ligand complex is calculated by first refining both components individually using intramolecular NOEs, and then docking the two using intermolecular NOEs.

For example, the ^{1}H, ^{15}N-HSQC spectrum for the fluorescein-binding monobody LB25.5 is shown in Figure 14. The spectrum shows a good dispersion (peaks are spread out) indicating that LB25.5 is folded into a globular conformation. Further, the spectrum resembles that for the wild-type Fn3, showing that the overall structure of LB25.5 is similar to that of Fn3. These results demonstrate that ligand-binding monobodies can be obtained without changing the global fold of the Fn3 scaffold.

Chemical shift perturbation experiments are performed by forming the complex between an isotope-labeled FnAb and an unlabeled ligand. The formation of a stoichiometric complex is followed by recording the HSQC spectrum. Because chemical shift is extremely sensitive to nuclear environment, formation of a complex usually results in substantial chemical shift changes for resonances of amino acid residues in the interface. Isotope-edited NMR experiments (2D HSQC and 3D CBCA(CO)NH) are used to identify the

resonances that are perturbed in the labeled component of the complex; *i.e.* the monobody. Although the possibility of artifacts due to long-range conformational changes must always be considered, substantial differences for residues clustered on continuous surfaces are most likely to arise from direct contacts (Chen *et al.*, 1993; Gronenborn & Clore, 1993).

An alternative method for mapping the interaction surface utilizes amide hydrogen exchange (HX) measurements. HX rates for each amide proton are measured for ^{15}N labeled monobody both free and complexed with a ligand. Ligand binding is expected to result in decreased amide HX rates for monobody residues in the interface between the two proteins, thus identifying the binding surface. HX rates for monobodies in the complex are measured by allowing HX to occur for a variable time following transfer of the complex to D_2O; the complex is dissociated by lowering pH and the HSQC spectrum is recorded at low pH where amide HX is slow. Fn3 is stable and soluble at low pH, satisfying the prerequisite for the experiments.

EXAMPLE XVI

Construction and Analysis of Fn3-Display System Specific for Ubiquitin

An Fn3-display system was designed and synthesized, ubiquitin-binding clones were isolated and a major Fn3 mutant in these clones was biophysically characterized.

Gene construction and phage display of Fn3 was performed as in Examples I and II above. The Fn3-phage pIII fusion protein was expressed from a phagemid-display vector, while the other components of the M13 phage, including the wild-type pIII, were produced using a helper phage (Bass *et al.*, 1990). Thus, a phage produced by this system should contain less than one copy of Fn3 displayed on the surface. The surface display of Fn3 on the phage was detected by ELISA using an anti-Fn3 antibody. Only phages containing the Fn3-pIII fusion vector reacted with the antibody.

After confirming the phage surface to display Fn3, a phage display library of Fn3 was constructed as in Example III. Random sequences were introduced in the BC and FG loops. In the first library, five residues (77-81) were

randomized and three residues (δ2-84) were deleted from the FG loop. The deletion was intended to reduce the flexibility and improve the binding affinity of the FG loop. Five residues (26-30) were also randomized in the BC loop in order to provide a larger contact surface with the target molecule. Thus, the resulting library contains five randomized residues in each of the BC and FG loops (Table 7). This library contained approximately 10^8 independent clones.

Library Screening

Library screening was performed using ubiquitin as the target molecule. In each round of panning, Fn3-phages were absorbed to a ubiquitin-coated surface, and bound phages were eluted competitively with soluble ubiquitin. The recovery ratio improved from 4.3×10^{-7} in the second round to 4.5×10^{-6} in the fifth round, suggesting an enrichment of binding clones. After five founds of panning, the amino acid sequences of individual clones were determined (Table 7).

Table 7. Sequences in the variegated loops of enriched clones

Name	BC loop	FG loop	Frequency
Wild Type	GCAGTTACCGTGCGT (SEQ ID NO:93) AlaValThrValArg (SEQ ID NO:94)	GGCCGTGGTGACAGCCCAGCGAGC (SEQ ID NO:95) GlyArgGlyAspSerProAlaSer (SEQ ID NO:96)	–
Library[a]	NNKNNKNNKNNKNNK X X X X X	NNKNNKNNKNNKNNK--------- X X X X X (deletion)	–
clone1 (Ubi4)	TCGAGGTTGCGGCGG (SEQ ID NO:97) SerArgLeuArgArg (SEQ ID NO:98)	CCGCCGTGGAGGGTG (SEQ ID NO:99) ProProTrpArgVal (SEQ ID NO:100)	9
clone2	GGTCAGCGAACTTTT (SEQ ID NO:101) GlyGlnArgThrPhe (SEQ ID NO:102)	AGGCGGTGGTGGGCT (SEQ ID NO:103) ArgArgTrpTrpAla (SEQ ID NO:104)	1

clone3	GCGAGGTGGACGCTT (SEQ ID NO:105) AlaArgTrpThrLeu (SEQ ID NO:106)	AGGCGGTGGTGGTGG (SEQ ID NO:107) ArgArgTrpTrpTrp (SEQ ID NO:108)	1

[a] N denotes an equimolar mixture of A, T, G and C; K denotes an equimolar mixture of G and T.

A clone, dubbed Ubi4, dominated the enriched pool of Fn3 variants. Therefore, further investigation was focused on this Ubi4 clone. Ubi4 contains four mutations in the BC loop (Arg 30 in the BC loop was conserved) and five mutations and three deletions in the FG loop. Thus 13% (12 out of 94) of the residues were altered in Ubi4 from the wild-type sequence.

Figure 15 shows a phage ELISA analysis of Ubi4. The Ubi4 phage binds to the target molecule, ubiquitin, with a significant affinity, while a phage displaying the wild-type Fn3 domain or a phase with no displayed molecules show little detectable binding to ubiquitin (Figure 15a). In addition, the Ubi4 phage showed a somewhat elevated level of background binding to the control surface lacking the ubiquitin coating. A competition ELISA experiments shows the IC_{50} (concentration of the free ligand which causes 50% inhibition of binding) of the binding reaction is approximately 5 μM (Fig. 15b). BSA, bovine ribonuclease A and cytochrome C show little inhibition of the Ubi4-ubiquitin binding reaction (Figure 15c), indicating that the binding reaction of Ubi4 to ubiquitin does result from specific binding.

Characterization of a Mutant Fn3 Protein

The expression system yielded 50-100 mg Fn3 protein per liter culture. A similar level of protein expression was observed for the Ubi4 clone and other mutant Fn3 proteins.

Ubi4-Fn3 was expressed as an independent protein. Though a majority of Ubi4 was expressed in E. coli as a soluble protein, its solubility was found to be significantly reduced as compared to that of wild-type Fn3. Ubi4 was soluble up to ~20 μM at low pH, with much lower solubility at neutral pH. This solubility was not high enough for detailed structural characterization using

NMR spectroscopy or X-ray crystallography.

The solubility of the Ubi4 protein was improved by adding a solubility tail, GKKGK (SEQ ID NO:109), as a C-terminal extension. The gene for Ubi4-Fn3 was subcloned into the expression vector pAS45 using PCR. The C-terminal solubilization tag, GKKGK (SEQ ID NO:109), was incorporated in this step. E. coli BL21 (DE3) (Novagen) was transformed with the expression vector (pAS45 and its derivatives). Cells were grown in M9 minimal media and M9 media supplemented with Bactotryptone (Difco) containing ampicillin (200 μg/ml). For isotopic labeling, ^{15}N NH_4Cl replaced unlabeled NH_4Cl in the media. 500 ml medium in a 2 liter baffle flask was inoculated with 10 ml of overnight culture and agitated at 37°C. IPTG was added at a final concentration of 1 mM to initiate protein expression when OD (600 nm) reaches one. The cells were harvested by centrifugation 3 hours after the addition of IPTG and kept frozen at -70°C until used.

Proteins were purified as follows. Cells were suspended in 5 ml/(g cell) of Tris (50 mM, pH 7.6) containing phenylmethylsulfonyl fluoride (1 mM). Hen egg lysozyme (Sigma) was added to a final concentration of 0.5 mg/ml. After incubating the solution for 30 minutes at 37°C, it was sonicated three times for 30 seconds on ice. Cell debris was removed by centrifugation. Concentrated sodium chloride was added to the solution to a final concentration of 0.5 M. The solution was applied to a Hi-Trap chelating column (Pharmacia) preloaded with nickel and equilibrated in the Tris buffer containing sodium chloride (0.5 M). After washing the column with the buffer, histag-Fn3 was eluted with the buffer containing 500 mM imidazole. The protein was further purified using a ResourceS column (Pharmacia) with a NaCl gradient in a sodium acetate buffer (20 mM, pH 4.6).

With the GKKGK (SEQ ID NO:109) tail, the solubility of the Ubi4 protein was increased to over 1 mM at low pH and up to ~50 μM at neutral pH. Therefore, further analyses were performed on Ubi4 with this C-terminal extension (hereafter referred to as Ubi4-K). It has been reported that the solubility of a minibody could be significantly improved by addition of three Lys residues at the N- or C-termini (Bianchi *et al.*, 1994). In the case of protein Rop,

a non-structured C-terminal tail is critical in maintaining its solubility (Smith *et al.*, 1995).

Oligomerization states of the Ubi4 protein were determined using a size exclusion column. The wild-type Fn3 protein was monomeric at low and neutral pH's. However, the peak of the Ubi4-K protein was significantly broader than that of wild-type Fn3, and eluted after the wild-type protein. This suggests interactions between Ubi4-K and the column material, precluding the use of size exclusion chromatography to determine the oligomerization state of Ubi4. NMR studies suggest that the protein is monomeric at low pH.

The Ubi4-K protein retained a binding affinity to ubiquitin as judged by ELISA (Figure 15d). However, an attempt to determine the dissociation constant using a biosensor (Affinity Sensors, Cambridge, U.K.) failed because of high background binding of Ubi4-K-Fn3 to the sensor matrix. This matrix mainly consists of dextran, consistent with the observation that interactions between Ubi4-K interacts with the cross-linked dextran of the size exclusion column.

Example XVII

Stability Measurements of Monobodies

Guanidine hydrochloride (GuHCl)-induced unfolding and refolding reactions were followed by measuring tryptophan fluorescence. Experiments were performed on a Spectronic AB-2 spectrofluorometer equipped with a motor-driven syringe (Hamilton Co.). The cuvette temperature was kept at 30°C. The spectrofluorometer and the syringe were controlled by a single computer using a home-built interface. This system automatically records a series of spectra following GuHCl titration. An experiment started with a 1.5 ml buffer solution containing 5 μM protein. An emission spectrum (300-400 nm; excitation at 290 nm) was recorded following a delay (3-5 minutes) after each injection (50 or 100 μl) of a buffer solution containing GuHCl. These steps were repeated until the solution volume reached the full capacity of a cuvette (3.0 ml). Fluorescence intensities were normalized as ratios to the intensity at an isofluorescent point which was determined in separate experiments. Unfolding curves were fitted with a two-state model using a nonlinear least-squares routine

(Santoro & Bolen, 1988). No significant differences were observed between experiments with delay times (between an injection and the start of spectrum acquisition) of 2 minutes and 10 minutes, indicating that the unfolding/refolding reactions reached close to an equilibrium at each concentration point within the delay times used.

Conformational stability of Ubi4-K was measured using above-described GuHCl-induced unfolding method. The measurements were performed under two sets of conditions; first at pH 3.3 in the presence of 300 mM sodium chloride, where Ubi4-K is highly soluble, and second in TBS, which was used for library screening. Under both conditions, the unfolding reaction was reversible, and we detected no signs of aggregation or irreversible unfolding. Figure 16 shows unfolding transitions of Ubi4-K and wild-type Fn3 with the N-terminal $(his)_6$ tag and the C-terminal solubility tag. The stability of wild-type Fn3 was not significantly affected by the addition of these tags. Parameters characterizing the unfolding transitions are listed in Table 8.

Table 8. Stability parameters for Ubi4 and wild-type Fn3 as determined by GuHCl-induced unfolding

Protein	ΔG_0 (kcal mol^{-1})	m_G (kcal mol^{-1} M^{-1})
Ubi4 (pH 7.5)	4.8 ± 0.1	2.12 ± 0.04
Ubi4 (pH 3.3)	6.5 ± 0.1	2.07 ± 0.02
Wild-type (pH 7.5)	7.2 ± 0.2	1.60 ± 0.04
Wild-type (pH 3.3)	11.2 ± 0.1	2.03 ± 0.02

ΔG_0 is the free energy of unfolding in the absence of denaturant; m_G is the dependence of the free energy of unfolding on GuHCl concentration. For solution conditions, see Figure 4 caption.

Though the introduced mutations in the two loops certainly decreased the stability of Ubi4-K relative to wild-type Fn3, the stability of Ubi4 remains comparable to that of a "typical" globular protein. It should also be noted that the stabilities of the wild-type and Ubi4-K proteins were higher at pH 3.3 than at pH 7.5.

The Ubi4 protein had a significantly reduced solubility as compared to

that of wild-type Fn3, but the solubility was improved by the addition of a solubility tail. Since the two mutated loops include the only differences between the wild-type and Ubi4 proteins, these loops must be the origin of the reduced solubility. At this point, it is not clear whether the aggregation of Ubi4-K is caused by interactions between the loops, or by interactions between the loops and the invariable regions of the Fn3 scaffold.

The Ubi4-K protein retained the global fold of Fn3, showing that this scaffold can accommodate a large number of mutations in the two loops tested. Though the stability of the Ubi4-K protein is significantly lower than that of the wild-type Fn3 protein, the Ubi4 protein still has a conformational stability comparable to those for small globular proteins. The use of a highly stable domain as a scaffold is clearly advantageous for introducing mutations without affecting the global fold of the scaffold. In addition, the GuHCl-induced unfolding of the Ubi4 protein is almost completely reversible. This allows the preparation of a correctly folded protein even when a Fn3 mutant is expressed in a misfolded form, as in inclusion bodies. The modest stability of Ubi4 in the conditions used for library screening indicates that Fn3 variants are folded on the phage surface. This suggests that a Fn3 clone is selected by its binding affinity in the folded form, not in a denatured form. Dickinson *et al.* proposed that Val 29 and Arg 30 in the BC loop stabilize Fn3. Val 29 makes contact with the hydrophobic core, and Arg 30 forms hydrogen bonds with Gly 52 and Val 75. In Ubi4-Fn3, Val 29 is replaced with Arg, while Arg 30 is conserved. The FG loop was also mutated in the library. This loop is flexible in the wild-type structure, and shows a large variation in length among human Fn3 domains (Main *et al.*, 1992). These observations suggest that mutations in the FG loop may have less impact on stability. In addition, the N-terminal tail of Fn3 is adjacent to the molecular surface formed by the BC and FG loops (Figure 1 and 17) and does not form a well-defined structure. Mutations in the N-terminal tail would not be expected to have strong detrimental effects on stability. Thus, residues in the N-terminal tail may be good sites for introducing additional mutations.

Example XVIII

NMR Spectroscopy of Ubi4-Fn3

Ubi4-Fn3 was dissolved in [^{2}H]-Gly HCl buffer (20 mM, pH 3.3) containing NaCl (300 mM) using an Amicon ultrafiltration unit. The final protein concentration was 1 mM. NMR experiments were performed on a Varian Unity INOVA 600 spectrometer equipped with a triple-resonance probe with pulsed field gradient. The probe temperature was set at 30°C. HSQC, TOCSY-HSQC and NOESY-HSQC spectra were recorded using published procedures (Kay *et al.*, 1992; Zhang *et al.*, 1994). NMR spectra were processed and analyzed using the NMRPipe and NMRView software (Johnson & Blevins, 1994; Delaglio *et al.*, 1995) on UNIX workstations. Sequence-specific resonance assignments were made using standard procedures (Wüthrich, 1986; Clore & Gronenborn, 1991). The assignments for wild-type Fn3 (Baron *et al.*, 1992) were confirmed using a ^{15}N-labeled protein dissolved in sodium acetate buffer (50 mM, pH 4.6) at 30°C.

The three-dimensional structure of Ubi4-K was characterized using this heteronuclear NMR spectroscopy method. A high quality spectrum could be collected on a 1 mM solution of ^{15}N-labeled Ubi4 (Figure 17a) at low pH. The linewidth of amide peaks of Ubi4-K was similar to that of wild-type Fn3, suggesting that Ubi4-K is monomeric under the conditions used. Complete assignments for backbone ^{1}H and ^{15}N nuclei were achieved using standard ^{1}H, ^{15}N double resonance techniques, except for a row of His residues in the N-terminal $(His)_6$ tag. There were a few weak peaks in the HSQC spectrum which appeared to originate from a minor species containing the N-terminal Met residue. Mass spectroscopy analysis showed that a majority of Ubi4-K does not contain the N-terminal Met residue. Fig. 17 shows differences in ^{1}HN and ^{15}N chemical shifts between Ubi4-K and wild-type Fn3. Only small differences are observed in the chemical shifts, except for those in and near the mutated BC and FG loops. These results clearly indicate that Ubi4-K retains the global fold of Fn3, despite the extensive mutations in the two loops. A few residues in the N-terminal region, which is close to the two mutated loops, also exhibit significant chemical differences between the two proteins. An HSQC spectrum was also recorded on a 50 μM sample of Ubi4-K in TBS. The spectrum was similar to

that collected at low pH, indicating that the global conformation of Ubi4 is maintained between pH 7.5 and 3.3.

Example XIX:

Stabilization of Fn3 domain by removing unfavorable electrostatic interactions on the protein surface

Introduction

Increasing the conformational stability of a protein by mutation is a major interest in protein design and biotechnology. The three-dimensional structures of proteins are stabilized by combination of different types of forces. The hydrophobic effect, van der Waals interactions and hydrogen bonds are known to contribute to stabilize the folded state of proteins (Kauzmann, W. (1959) *Adv. Prot. Chem. 14*, 1-63; Dill, K. A. (1990) *Biochemistry 29*, 7133-7155; Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996) *Faseb J 10*, 75-83). These stabilizing forces primarily originate from residues that are well packed in a protein, such as those that constitute the hydrophobic core. Because a change in the protein core would induce a rearrangement of adjacent moieties, it is difficult to improve protein stability by increasing these forces without massive computation (Malakauskas, S. M. & Mayo, S. L. (1998) *Nat Struct Biol 5*, 470-475). Ion pairs between charged groups are commonly found on the protein surface (Creighton, T. E. (1993) *Proteins: structures and molecular properties*, Freeman, New York), and an ion pair could be introduced to a protein with small structural perturbations. However, a number of studies have demonstrated that the introduction of an attractive electrostatic interaction, such as an ion pair, on protein surface has small effects on stability (Dao-pin, S., Sauer, U., Nicholson, H. & Matthews, B. W. (1991) *Biochemistry 30*, 7142-7153; Sali, D., Bycroft, M. & Fersht, A. R. (1991) *J. Mol. Biol. 220*, 779-788). A large desolvation penalty and the loss of conformational entropy of amino acid side chains oppose the favorable electrostatic contribution (Yang, A.-S. & Honig, B. (1992) *Curr. Opin. Struct. Biol. 2*, 40-45; Hendsch, Z. S. & Tidor, B. (1994) *Protein Sci. 3*, 211-226). Recent studies demonstrated that repulsive electrostatic interactions on the

protein surface, in contrast, may significantly destabilize a protein, and that it is possible to improve protein stability by optimizing surface electrostatic interactions (Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999) *Biochemistry 38*, 16419-16423; Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000) *Nat Struct Biol 7*, 380-383; Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B. & Raleigh, D. P. (2000) *Biochemistry 39*, 872-879; Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999) *Protein Sci 8*, 1843-1849). In the present experiments, the inventor improved protein stability by modifying surface electrostatic interactions.

During the characterization of monobodies it was found that these proteins, as well as wild-type FNfn10, are significantly more stable at low pH than at neutral pH (Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151). These observations indicate that changes in the ionization state of some moieties in FNfn10 modulate the conformational stability of the protein, and suggest that it might be possible to enhance the conformational stability of FNfn10 at neutral pH by adjusting electrostatic properties of the protein. Improving the conformational stability of FNfn10 will also have practical importance in the use of FNfn10 as a scaffold in biotechnology applications.

Described below are experiments that detailed characterization of the pH dependence of FNfn10 stability, identified unfavorable interactions between side chain carboxyl groups, and improved the conformational stability of FNfn10 by point mutations on the surface. The results demonstrate that the surface electrostatic interactions contribute significantly to protein stability, and that it is possible to enhance protein stability by rationally modulating these interactions.

Experimental Procedures

Protein expression and purification

The wild-type protein used for the NMR studies contained residues 1-94 of FNfn10 (residue numbering is according to Figure 2(a) of Koide et al. (Koide,

A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151)), and additional two residues (Met-Gln) at the N-terminus (these two residues are numbered -2 and -1, respectively). The gene coding for the protein was inserted in pET3a (Novagen, WI). *Eschericha coli* BL21 (DE3) transformed with the expression vector was grown in the M9 minimal media supplemented with ^{13}C-glucose and ^{15}N-ammonium chloride (Cambridge Isotopes) as the sole carbon and nitrogen sources, respectively. Protein expression was induced as described previously (Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151). After harvesting the cells by centrifuge, the cells were lysed as described (Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151). After centrifugation, supernatant was dialyzed against 10 mM sodium acetate buffer (pH 5.0), and the protein solution was applied to a SP-Sepharose FastFlow column (Amersham Pharmacia Biotech), and FN3 was eluted with a gradient of sodium chloride. The protein was concentrated using an Amicon concentrator using YM-3 membrane (Millipore).

The wild-type protein used for the stability measurements contained an N-terminal histag (MGSSHHHHHHSSGLVPRGSH) (SEQ ID NO:114) and residues -2-94 of FNfn10. The gene for FN3 described above was inserted in pET15b (Novagen). The protein was expressed and purified as described (Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151). The wild-type protein used for measurements of the pH dependence shown in Figure 22 contained Arg 6 to Thr mutation, which had originally been introduced to remove a secondary thrombin cleavage site (Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151). Because Asp 7, which is adjacent to Arg 6, was found to be critical in the pH dependence of FN3 stability as detailed under Results, subsequent studies were performed using the wild-type, Arg 6, background. The genes for the D7N and D7K mutants were constructed using standard polymerase chain reactions, and inserted in pET15b. These proteins were prepared in the same manner as for the wild-type protein. ^{13}C, ^{15}N-labeled proteins for pK_a measurements were prepared as described above, and the histag moiety was not removed from these proteins.

Chemical denaturation measurements

Proteins were dissolved to a final concentration of 5 μM in 10 mM sodium citrate buffer at various pH containing 100 mM sodium chloride. Guanidine HCl (GuHCl)-induce unfolding experiments were performed as described previously (Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151; Koide, S., Bu, Z., Risal, D., Pham, T.-N., Nakagawa, T., Tamura, A. & Engelman, D. M. (1999) *Biochemistry 38*, 4757-4767). GuHCl concentration was determined using an Abbe refractometer (Spectronic Instruments) as described (Pace, C. N. & Sholtz, J. M. (1997) in *Protein structure. A practical approach* (Creighton, T. E., Ed.) Vol. pp299-321, IRL Press, Oxford). Data were analyzed according to the two-state model as described (Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151; Santoro, M. M. & Bolen, D. W. (1988) *Biochemistry 27*, 8063-8068.).

Thermal denaturation measurements

Proteins were dissolved to a final concentration of 5 μM in 20 mM sodium phosphate buffer (pH 7.0) containing 0.1 or 1 M sodium chloride or in 20 mM glycine HCl buffer (pH 2.4) containing 0.1 or 1 M sodium chloride. Additionally 6.3 M urea was included in all solutions to ensure reversibility of the thermal denaturation reaction. In the absence of urea it was found that denatured FNfn10 adheres to quartz surface, and that the thermal denaturation reaction was irreversible. Circular dichroism measurements were performed using a Model 202 spectrometer equipped with a Peltier temperature controller (Aviv Instruments). A cuvette with a 0.5-cm pathlength was used. The ellipticity at 227 nm was recorded as the sample temperature was raised at a rate of approximately 1 °C per minute. Because of decomposition of urea at high temperature, the pH of protein solutions tended to shift upward during an experiment. The pH of protein solution was measured before and after each thermal denaturation measurement to ensure that a shift no more than 0.2 pH unit occurred in each measurement. At pH 2.4, two sections of a thermal denaturation curve (30-65 °C and 60-95 °C) were acquired from separate

samples, in order to avoid a large pH shift. The thermal denaturation data were fit with the standard two-state model (Pace, C. N. & Sholtz, J. M. (1997) in *Protein structure. A practical approach* (Creighton, T. E., Ed.) Vol. pp299-321, IRL Press, Oxford):

$$\Delta G(T) = \Delta H_m (1 - T/T_m) - \Delta C_P[(T_m - T) + T \ln(T/T_m)]$$

where $\Delta G(T)$ is the Gibbs free energy of unfolding at temperature T, ΔH_m is the enthalpy change upon unfolding at the midpoint of the transition, T_m, and ΔC_P is the heat capacity change upon unfolding. The value for ΔC_P was fixed at 1.74 kcal mol^{-1} K^{-1}, according to the approximation of Myers et al. (Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995) *Protein Sci. 4*, 2138-2148). Most of the datasets taken in the presence of 1 M NaCl did not have a sufficient baseline for the unfolded state, and thus it was assumed the slope of the unfolded baseline in the presence of 1 M NaCl to be identical to that determined in the presence of 0.1 M NaCl.

NMR spectroscopy

NMR experiments were performed at 30 °C on an INOVA 600 spectrometer (Varian Instruments). The C(CO)NH experiment (Grzesiek, S., Anglister, J. & Bax, A. (1993) *J. Magn. Reson. B 101*, 114-119) and the CBCACOHA experiment (Kay, L. E. (1993) *J. Am. Chem. Soc. 115*, 2055-2057) were collected on a [^{13}C, ^{15}N]-wild-type FNfn10 sample (1 mM) dissolved in 50 mM sodium acetate buffer (pH 4.6) containing 5 % (v/v) deuterium oxide, using a Varian 5 mm triple resonance probe with pulsed field gradient. The carboxyl ^{13}C resonances were assigned based on the backbone ^{1}H, ^{13}C and ^{15}N resonance assignments of FNfn10 (Baron, M., Main, A. L., Driscoll, P. C., Mardon, H. J., Boyd, J. & Campbell, I. D. (1992) *Biochemistry 31*, 2068-2073). pH titration of carboxyl resonances were performed on a 0.3 mM FNfn10 sample dissolved in 10 mM sodium citrate containing 100 mM sodium chloride and 5 % (v/v) deuterium oxide. An 8 mm triple-resonance, pulse-field gradient probe (Nanolac Corporation) was used for pH titration. Two-dimensional H(C)CO

spectra were collected using the CBCACOHA pulse sequence as described previously (McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Koerner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. & Withers, S. G. (1996) *Biochemistry 35*, 9958-9966). Sample pH was changed by adding small aliquots of hydrochloric acid, and pH was measured before and after taking NMR data. ^{1}H, ^{15}N-HSQC spectra were taken as described previously (Kay, L. E., Keifer, P. & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663-10665). NMR data were processed using the NMRPipe package (Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995) *J. Biomol. NMR 6*, 277-293), and analyzed using the NMRView software (Johnson, B. A. & Blevins, R. A. (1994) *J. Biomol. NMR 4*, 603-614).

NMR titration curves of the carboxyl ^{13}C resonances were fit to the Henderson-Hasselbalch equation to determine pK_a's:

$$\delta(pH) = (\delta_{acid} + \delta_{base} 10^{(pH-pK_a)}) / (1 + 10^{(pH-pKa)})$$

where δ is the measured chemical shift, δ_{acid} is the chemical shift associated with the protonated state, δ_{base} is the chemical shift associated with the deprotonated state, and pK_a is the pK_a value for the residue. Data were also fit to an equation with two ionizable groups:

$$\delta(pH) = (\delta_{AH2} + \delta_{AH} 10^{(pH-pK_{a1})} + \delta_{A} 10^{(2pH-pK_{a1}-pK_{a2})}) / (1 + 10^{(pH-pK_{a1})} + 10^{(2pH-pK_{a1}-pK_{a2})})$$

where δ_{AH2}, δ_{AH} and δ_{A} are the chemical shifts associated with the fully protonated, singularly protonated and deprotonated states, respectively, and pK_{a1} and pK_{a2} are pK_a's associated with the two ionization steps. Data fitting was performed using the nonlinear least-square regression method in the program Igor Pro (WaveMetrix, OR) on a Macintosh computer.

Results

pH Dependence of FNfn10 stability

Previously, it was found that FNfn10 is more stable at acidic pH than at neutral pH (Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151). In the present experiments, the pH dependence of its stability was further characterized. Because of its high stability, FNfn10 could not be fully denatured in urea at 30 °C. Thus GuHCl-induced chemical denaturation (Figure 18) was used. The denaturation reaction was fully reversible under all conditions tested. In order to minimize errors caused by extrapolation, the free energy of unfolding at 4 M GuHCl was used for comparison (Figure 18). The stability increased as the pH was lowered, with apparent plateaus at both ends of the pH range. The pH dependence curve has an apparent transition midpoint near pH 4. In addition, a gradual increase in the *m* value, the dependence of the unfolding free energy on denaturant concentration was noted. Pace *et al.* reported a similar pH dependence of the *m* value for barnase (Pace, C. N., Laurents, D. V. & Erickson, R. E. (1992) *Biochemistry 31*, 2728-2734). These results indicate that FNfn10 contains interactions that stabilize the protein at low pH, or those that destabilize it at neutral pH. The results also suggest that by identifying and altering the interactions that give rise to the pH dependence, one may be able to improve the stability of FNfn10 at neutral pH to a degree similar to that found at low pH.

Determination of pK_a's of the side chain carboxyl groups in wild-type FNfn10

The pH dependence of FNfn10 stability suggests that amino acids with pK_a near 4 are involved in the observed transition. The carboxyl groups of Asp and Glu generally have pK_a in this range (Creighton, T. E. (1993) *Proteins: structures and molecular properties*, Freeman, New York). It is well known that if a carboxyl group has unfavorable (i.e. destabilizing) interactions in the folded state, its pK_a is shifted to a higher value from its unperturbed value (Yang, A.-S. & Honig, B. (1992) *Curr. Opin. Struct. Biol. 2*, 40-45). If a carboxyl group has favorable interactions in the folded state, it has a lower pK_a. Thus, the pK_a values of all carboxylates in FNfn10 using heteronuclear NMR spectroscopy were determined in order to identify stabilizing and destabilizing interactions involving carboxyl groups.

First, the ^{13}C resonance for the carboxyl carbon of each Asp and Glu residue in FN3 was assigned (Figure 19). Next, pH titration of the ^{13}C resonances for these groups was performed (Figure 20). Titration curves for Asp 3, 67 and 80, and Glu 38 and 47 could be fit well with the Henderson-Hasselbalch equation with a single pK_a. The pK_a values for these residues (Table 9) are either close to or slightly lower than their respective unperturbed values (3.8-4.1 for Asp, and 4.1-4.6 for Glu (Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. (1999) *Biochemistry 38*, 4896-4903)), indicating that these carboxyl groups are involved in neutral or slightly favorable electrostatic interactions in the folded state.

Table 9. pK_a values for Asp and Glu residues in FN3[1].

Residue	Protein		
	Wild-Type	D7N	D7K
E9	3.84, 5.40[2]	4.98	4.53
E38	3.79	3.87	3.86
E47	3.94	3.99	3.99
D3	3.66	3.72	3.74
D7	3.54, 5.54[2]	-	-
D23	3.54, 5.25[2]	3.68	3.82
D67	4.18	4.17	4.14
D80	3.40	3.49	3.48

[1]The standard deviations in the pK_a values are less than 0.05 pH units for those fit with a single pK_a and less than 0.15 pH unit for those with two pK_a's.

[2]Data for E9, D7 and D23 were fit with a transition curve with two pK_a values.

The titration curves for Asp 7 and 23, and Glu 9 were fit better with the Henderson-Hasselbalch equation with two pK_a values, and one of the two pK_a values for each were shifted higher than the respective unperturbed values (Figure 19B). The titration curves with two apparent pK_a values of these carboxyl groups may be due to influence of an ionizable group in the vicinity. In the three-dimensional structure of FNfn10 (Main, A. L., Harvey, T. S., Baron, M., Boyd, J.

& Campbell, I. D. (1992) *Cell 71*, 671-678), Asp 7 and 23, and Glu 9 form a patch on the surface (Figure 21), with Asp 7 centrally located in the patch. Thus, it is reasonable to expect that these residues influence each other's ionization profile. In order to identify which of the three residues have a highly upshifted pK_a, the H(C)CO spectrum of the protein in 99 % D_2O buffer at pH* 5.0 (direct pH meter reading) was then collected. Asp 23 and Glu 9 showed larger deuterium isotope shifts (0.33 and 0.32 ppm, respectively) than Asp 7 (0.18 ppm). These results show that Asp 23 and Glu 9 are protonated to a greater degree than Asp 7. Thus, we concluded that Asp 23 and Glu 9 have highly upshifted pK_a's, due to strong influence of Asp 7.

Mutational analysis

The spatial proximity of Asp 7 and 23, and Glu 9 explains the unfavorable electrostatic interactions in FNfn10 identified in this study. At low pH where these residues are protonated and neutral, the repulsive interactions are expected to be mostly relieved. Thus, it should be possible to improve the stability of FNfn10 at neutral pH, by removing the electrostatic repulsion between these three residues. Because Asp 7 is centrally located among the three residues, it was decided to mutate Asp 7. Two mutants, D7N and D7K were prepared. The former neutralizes the negative charge with a residue of virtually identical size. The latter places a positive charge at residue 7 and increases the size of the side chain.

The ^{1}H, ^{15}N-HSQC spectra of the two mutant proteins were nearly identical to that of the wild-type protein, indicating that these mutations did not cause large structural perturbations (data not shown). The degrees of stability of the mutant proteins were then characterized using thermal and chemical denaturation measurements. Thermal denaturation measurements were performed initially with 100 mM sodium chloride, and 6.3 M urea was included to ensure reversible denaturation and to decrease the temperature of the thermal transition. All the proteins were predominantly folded in 6.3 M urea at room temperature. All the proteins underwent a cooperative transition, and the two mutants were found to be significantly more stable than the wild type at neutral

pH (Figure 22 and Table 10). Furthermore, these mutations almost eliminated the pH dependence of the conformational stability of FNfn10. These results confirmed that destabilizing interactions involving Asp 7 in wild-type FNfn10 at neutral pH are the primary cause of the pH dependence.

Table 10. The midpoint of thermal denaturation (in °C) of wild-type and mutant FN3 in the presence of 6.3 M urea.

Protein	pH 2.4		pH 7.0	
	0.1 M NaCl	1M NaCl	0.1 M NaCl	1 M NaCl
wild type	72	82	62	70
D7N	68	82	69	80
D7K	69	77	70	78

The error in the midpoints for the 0.1 M NaCl data is ± 0.5 °C. Because most of the 1M NaCl data did not have a sufficient baseline for the denatured state, the error in the midpoints for these data was estimated to be ±2 °C.

The effect of increased sodium chloride concentration on the conformational stability of the wild type and the two mutant proteins was next investigated. All proteins were more stable in 1 M sodium chloride than in 0.1 M sodium chloride (Figure 22). The increase of the sodium chloride concentration elevated the T_m of the mutant proteins by approximately 10 °C at both acidic and neutral pH (Table 10). Remarkably the wild-type protein was also equally stabilized at both pH, although it contains unfavorable interactions among the carboxyl groups at neutral pH but not at acidic pH.

Chemical denaturation of FNfn10 proteins was monitored using fluorescence emission from the single Trp residue of FNfn10 (Figure 23). The free energies of unfolding at pH 6.0 and 4 M GuHCl were determined to be 1.1 (± 0.3), 1.7 (± 0.2) and 1.4 (± 0.1) kcal/mol for the wild type, D7N and D7K, respectively, indicating that the two mutations also increased the conformational stability against chemical denaturation.

Determination of the pK_a's of the side chain carboxyl groups in the mutant proteins

The ionization properties of carboxyl groups in the two mutant proteins was investigated. The 2D H(C)CO spectra of the mutant proteins at the high and low ends of the pH titration (pH ~7 and ~1.5, respectively) were nearly identical to the respective spectra of the wild type, except for the loss of the cross peaks for Asp 7 (data not shown). This similarity allowed for an unambiguous assignment of resonances of the mutants, based on the assignments for wild-type FNfn10. The pH titration experiments revealed that, except for Glu 9 and Asp 23, the behaviors of Asp and Glu carboxyl groups are very close to their counterparts in the wild-type protein (Figure 24 Panels A, C, D, F and G, and Table 9), indicating that the two mutations have marginal effects on the electrostatic environments for these carboxylates. In contrast, the titration curves for E9 and D23 show significant changes upon mutation (Figure 24 Panels B and E). The pK_a of D23 was lowered by more than 1.6 and 1.4 pH units in the D7N and D7K mutants, respectively. These results clearly show that the repulsive interaction between D7 and D23 contributes to the increase in pK_a of Asp 23 in the wild-type protein, and that it was eliminated by the neutralization of the negative charge at residue 7. The pK_a of Glu 9 was reduced by 0.4 pH unit by the D7N mutation, while it was decreased by 0.8 pH units in the D7K mutant. The greater reduction of Glu 9 pK_a by the D7K mutation suggests that there is a favorable interaction between Lys 7 and Glu 9 in this mutant protein.

Discussion

The present inventor has identified unfavorable electrostatic interactions in FNfn10, and improved its conformational stability by mutations on the protein surface. The results demonstrate that repulsive interactions between like charges on protein surface significantly destabilize a protein. The results are also consistent with recent reports by other groups (Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999) *Biochemistry 38*, 16419-16423; Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000) *Nat Struct Biol 7*, 380-383; Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S.,

Fairman, R., Tidor, B. & Raleigh, D. P. (2000) *Biochemistry 39*, 872-879; Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999) *Protein Sci 8*, 1843-1849), in which protein stability was improved by eliminating unfavorable electrostatic interactions on the surface. In these studies, candidates for mutations were identified by electrostatic calculations (Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999) *Biochemistry 38*, 16419-16423; Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B. & Raleigh, D. P. (2000) *Biochemistry 39*, 872-879; Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999) *Protein Sci 8*, 1843-1849) or by sequence comparison of homologous proteins with different stability (Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000) *Nat Struct Biol 7*, 380-383). The present strategy using pK_a determination using NMR has both advantages and disadvantages over the other strategies. The present method directly identifies residues that destabilize a protein. Also it does not depend on the availability of the high-resolution structure of the protein of interest. Electrostatic calculations may have large errors due to the flexibility of amino acid side chains on the surface, and the uncertainty in the dielectric constant on the protein surface and in the protein interior. For example, in the NMR structure of FNfn10 (Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992) *Cell 71*, 671-678), the root mean squared deviations among 16 model structures for the O^{ϵ} atom of Glu residues are 1.2-2.4 Å, and those for Lys N^{ζ} atoms are 1.5-3.1 Å. Such uncertainties in atom position can potentially cause large differences in calculation results. On the other hand, the present strategy requires the NMR assignments for carboxyl residues, and NMR measurements over a wide pH range. Although recent advances in NMR spectroscopy have made it straightforward to obtain resonance assignments for a small protein, some proteins may not be sufficiently soluble over the desired pH range. In addition, knowledge of the pK_a values of ionizable groups in the denatured state is necessary for accurately evaluating contributions of individual residues to stability (Yang, A.-S. & Honig, B. (1992) *Curr. Opin. Struct. Biol. 2*, 40-45).

Kuhlman *et al.* (Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. (1999) *Biochemistry 38*, 4896-4903) showed that pK_a's of carboxylates in the denatured state has a considerably large range than those obtained from small model compounds. Despite these limitations, the present method is applicable to many proteins.

The inventor showed that the unfavorable interactions involving the carboxyl groups of Asp 7, Glu 9 and Asp23 were no longer present if these groups are protonated at low pH or if Asp 7 was replaced with Asn or Lys. The similarity in the measured stability of the mutants and the wild type at low pH (Table 10) suggests that no other factors significantly contribute to the pH dependence of FNfn10 stability and that the mutations caused minimal structural perturbations. The little structural perturbation was expected, since the carboxyl groups of these three residues are at least 50 % exposed to the solvent, based on the solvent accessible surface area calculation on the NMR structure (Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992) *Cell 71*, 671-678).

The difference in thermal stability of the wild-type protein between acidic and neutral pH persisted in 1 M sodium chloride (Table 10). Likewise, the wild-type protein exhibited a large pH-dependence in stability in 4 M GuHCl (Figure 18). Furthermore, upon the increase in the sodium chloride concentration from 0.1 to 1.0 M, the T_m of the wild-type and mutant proteins all increased by ~10 °C, which is in the same magnitude as the change in T_m of the wild type by the pH shift. These data indicate that the unfavorable interactions identified in this study were not effectively shielded in 1 M NaCl or in 4 M GuHCl. Because the effect of increased sodium chloride was uniform, this stabilization effect of sodium chloride is likely due to the nonspecific salting-out effect (Timasheff, S. N. (1992) *Curr. Op. Struct. Biol. 2*, 35-39). Other groups also reported little shielding effect of salts on electrostatic interactions (Perutz, M. F., Gronenborn, A. M., Clore, G. M., Fogg, J. H. & Shih, D. T. (1985) *J Mol Biol 183*, 491-498; Hendsch, Z. S., Jonsson, T., Sauer, R. T. & Tidor, B. (1996) *Biochemistry 35*, 7621-7625). Electrostatic interactions are often thought to diminish with increasing ionic strength, particularly if the site of interaction is highly exposed. Accordingly, the present data at neutral pH (Table 10) showing no difference in

the salt sensitivity between the wild type and the mutants could be interpreted as Asp 7 not being responsible for destabilizing electrostatic interactions. Although the reason for this salt insensitivity is not yet clear, the present results provide a cautionary note on concluding the presence and absence of electrostatic interactions solely based on salt concentration dependence.

The carboxyl triad (Asp 7 and 23, and Glu 9) is highly conserved in FNfn10 from nine different organisms that were available in the protein sequence databank at National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). In these FNfn10 sequences, Asp 9 is conserved except one case where it is replaced with Asn, and Glu 9 is completely conserved. The position 23 is either Asp or Glu, preserving the negative charge. As was discovered in this study, the interactions among these residues are destabilizing. Thus, their high conservation, despite their negative effects on stability, suggests that these residues have functional importance in the biology of fibronectin. In the structure of a four-FN3 segment of human fibronectin (Leahy, D. J., Aukhil, I. & Erickson, H. P. (1996) *Cell 84*, 155-164), these residues are not directly involved in interactions with adjacent domains. Also these residues are located on the opposite face of FNfn10 from the integrin-binding RGD sequence in the FG loop (Figure 21). Therefore, it is not clear why these destabilizing residues are almost completely conserved in FNfn10. In contrast, no other FN3 domains in human fibronectin contain this carboxyl triad (for a sequence alignment, see ref Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992) *Cell 71*, 671-678). The carboxyl triad of FNfn10 may be involved in important interactions that have not been identified to date.

Clarke et al. (Clarke, J., Hamill, S. J. & Johnson, C. M. (1997) *J Mol Biol 270*, 771-778) reported that the stability of the third FN3 of human tenascin (TNfn3) increases as pH was decreased from 7 to 5. Although they could not perform stability measurements below pH 5 due to protein aggregation, the pH dependence of TNfn3 resembles that of FNfn10 shown in Figure 18. TNfn3 does not contain the carboxylate triad at positions 7, 9 and 23 (Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. (1992) *Science 258*, 987-991), indicating that the destabilization of TNfn3 at neutral pH is caused by a different

mechanism from that for FNfn10. A visual inspection of the TNfn3 structure revealed that it has a large number of carboxyl groups, and that Glu 834 and Asp 850 (numbering according to ref Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. (1992) *Science 258*, 987-991) forms a cross-strand pair. It will be interesting to examine whether altering this pair can increase the stability of TNfn3.

In conclusion, a strategy has been described to experimentally identify unfavorable electrostatic interactions on the protein surface and improve the protein stability by relieving such interactions. The present results have demonstrated that forming a repulsive interaction between carboxyl groups significantly destabilize a protein. This is in contrast to the small contributions of forming a solvent-exposed ion pair. Unfavorable electrostatic interactions on the surface seem quite common in natural proteins. Therefore, optimization of the surface electrostatic properties provides a generally applicable strategy for increasing protein stability (Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999) *Biochemistry 38*, 16419-16423; Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000) *Nat Struct Biol 7*, 380-383; Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B. & Raleigh, D. P. (2000) *Biochemistry 39*, 872-879; Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999) *Protein Sci 8*, 1843-1849). In addition, repulsive interactions between carboxylates can be exploited for destabilizing undesirable, alternate conformations in protein design ("negative design").

EXAMPLE XX

An extension of the carboxyl-terminus of the monobody scaffold

The wild-type protein used for stability measurements is described under Example 19. The carboxyl-terminus of the monobody scaffold was extended by four amino acid residues, namely, amino acid residues (Glu-Ile-Asp-Lys) (SEQ ID NO:119), which are the ones that immediately follow FNfn10 of human fibronectin. The extension was introduced into the FNfn10 gene using standard PCR methods. Stability measurements were performed as described under

Example 19. The free energy of unfolding of the extended protein was 7.4 kcal mol^{-1} at pH 6.0 and 30 °C, very close to that of the wild-type protein (7.7 kcal mol^{-1}). These results demonstrate that the C-terminus of the monobody scaffold can be extended without decreasing its stability.

The complete disclosure of all patents, patent documents and publications cited herein are incorporated by reference as if individually incorporated. The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described for variations obvious to one skilled in the art will be included within the invention defined by the claims.

REFERENCES

Alzari, P.N., Lascombe, M.-B. & Poljak, R.J. (1988) Three-dimensional structure of antibodies. *Annu. Rev. Immunol. 6*, 555-580.

Archer, S. J., Ikura, M., Torchia, D. A. & Bax, A. (1991) An alternative 3D NMR technique for correlating backbone 15N with side chain Hb resonances in large proteins *J. Magn. Reson. 95*, 636-641.

Aukhil, I., Joshi, P., Yan, Y. & Erickson, H. P. (1993) Cell- and heparin-binding domains of the hexabrachion arm identified by tenascin expression protein *J. Biol. Chem. 268*, 2542-2553.

Barbas, C. F., III, Kang, A. S., Lerner, R. A. & Benkovic, S. J. (1991) Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc. Natl. Acad. Sci. USA 88*, 7978-7982.

Barbas, C. F., III, Bain, J. D., Hoekstra, D. M. & Lerner, R. A. (1992) Semisynthetic combinatorial libraries: A chemical solution to the diversity problem *Proc. Natl. Acad. Sci. USA 89*, 4457-4461.

Baron, M., Main, A. L., Driscoll, P. C., Mardon, H. J., Boyd, J. & Campbell, I.D. (1992) ^{1}H NMR assignment and secondary structure of the cell adhesion type II module of fibronectin *Biochemistry 31*, 2068-2073.

Baron, M., Norman, D. G. & Campbell, I. D. (1991) Protein modules *Trends*

Biochem. Sci. 16, 13-17.

Bass, S., Greene, R. & Wells, J. A. (1990) Hormone phage: An enrichment method for variant proteins with altered binding properties *Proteins: Struct. Funct. Genet. 8*, 309-314.

Bax, A. & Grzesiek, S. (1993) Methodological advances in protein NMR. *Acc. Chem. Res. 26*, 131-138.

Becktel, W. J. & Schellman, J. A. (1987) Protein stability curves. *Biopolymer 26*, 1859-1877.

Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'acqua, W., Souchon, H., Schwarz, F. P., Mariuzza, R. A. & Poljak, R. J. (1994) Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA 91*, 1089-1093.

Bianchi, E., Venturini, S., Pessi, A., Tramontano, A. & Sollazzo, M. (1994) High level expression and rational mutagenesis of a designed protein, the minibody. From an insoluble to a soluble molecule. *J. Mol. Biol. 236*, 649.659.

Billeter, M., Neri, D., Otting, G., Qian, Y. Q. & Wüthrich, K. (1992) Precise vicinal coupling constants 3JHNa in proteins from nonlinear fits of J-modulated [^{15}N, ^{1}H]-COSY experiments. *J. Biomol. NMR 2*, 257-274.

Bodenhausen, G. & Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett. 69*, 185-189.

Bork, P. & Doolittle, R. F., *PNAS* 89:8990-8994 (1992).

Bork, P. & Doolittle, R. F. (1992) Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA 89*, 8990-8994.

Bork, P., Hom, L. & Sander, C. (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* 242, 309-320.

Brünger, A. T. (1992) *X-PLOR (Version 3.1): A system for X-ray crystallography and NMR.*, Yale Univ. Press, New Haven.

Burke, T., Bolger, R., Checovich, W. & Lowery, R. (1996) in *Phage display of peptides and proteins* (Kay, B. K., Winter, J. and McCafferty, J., Ed.) Vol. pp305-326, Academic Press, San Diego.

Campbell, I. D. & Spitzfaden, C. (1994) Building proteins with fibronectin type III modules *Structure 2*, 233-337.

Chen, Y., Reizer, J., Saier, M. H., Fairbrother, W. J. & Wright, P. E. (1993) Mapping the binding interfaces of the proteins of the bacterial phaphotransferase system, HPr and IIAglc. *Biochemistry 32*, 32-37.

Clarke, J., Hamill, S. J. & Johnson, C. M. (1997) *J Mol Biol 270*, 771-778.

Clackson & Wells, (1994) *Trends Biotechnology 12*, 173-184.

Clore, G. M. & Gronenborn, A. M. (1991) Structure of larger proteins in solution: Three- and four-dimensional heteronuclear NMR spectroscopy. *Science 252*, 1390-1399.

Corey, D. R., Shiau, A. K., Q., Y., Janowski, B. A. & Craik, C. S. (1993) Trypsin display on the surface of bacteriophage. *Gene 128*, 129-134.

Cota, E. & Clarke, J. (2000) *Protein Sci 9*, 112-120.

Creighton, T. E. (1993) *Proteins: structures and molecular properties*, Freeman, New York.

Dao-pin, S., Sauer, U., Nicholson, H. & Matthews, B. W. (1991) *Biochemistry 30*, 7142-7153.

Davies, J. & Riechmann, L. (1996). Single antibody domains as small recognition units: design and *in vitro* antigen selection of camelized, human VH domains with improved protein stability. Protein Eng., 9(6), 531-537.

Davies, J. & Riechmann, L. (1995) Antobody VH domains as small recognition units. *Bio/Technol. 13*, 475-479.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277-293.

Deng, W. P. & Nickoloff, J. A. (1992) Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Anal. Biochem. 200*, 81-88.

deVos, A. M., Ultsch, M. & Kossiakoff, A. A. (1992) Human Growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science 255*, 306-312.

Dickinson, C. D., Veerapandian, B., Dai, X.P., Hamlin, R. C., Xuong, N.-H.,

Ruoslahti, E. & Ely, K. R. (1994) Crystal structure of the tenth type III cell adhesion module of human fibronectin *J. Mol. Biol. 236*, 1079-1092..

Dill, K. A. (1990) *Biochemistry 29*, 7133-7155.

Djavadi-Ohaniance, L., Goldberg, M. E. & Friguet, B. (1996) in *Antibody Engineering. A Practical Approach* (McCafferty, J., Hoogenboom, H. R. and Chiswell, D. J., Ed.) Vol. pp. 77-97, Oxford Univ. Press, Oxford.

Dougall, W. C., Peterson, N. C. & Greene, M. I. (1994) Antibody-structure-based design of pharmacological agents. *Trends Biotechnol. 12*, 372-379.

Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991) A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson. 95*, 214-220.

Ghosh, G., Van Duyne, G., Ghosh, S. & Sigler, P. B. (1995) Structure of NF-kB p50 homodimer bound to a kB site. *Nature 373*, 303-310.

Gribskov, M., Devereux, J. & Burgess, R. R. (1984) The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. *Nuc. Acids. Res. 12*, 539-549.

Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999) *Protein Sci 8*, 1843-1849.

Groneborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991) A novel, highly stable fold of the immunoglobulin binding domain of Streptococcal protein G. *Science 253*, 657-661.

Gronenborn, A. M. & Clore, G. M. (1993) Identification of the contact surface of a Streptococcal protein G domain complexed with a human Fc fragment. *J. Mol. Biol. 233*, 331-335.

Grzesiek, S., Anglister, J. & Bax, A. (1993) Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. *J. Magn. Reson. B 101*, 114-119.

Grzesiek, S. & Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J.*

Am. Chem. Soc. 114, 6291-6293.

Grzesiek, S. & Bax, A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J. Biomol. NMR 3*, 185-204.

Harlow, E. & Lane, D. (1988) *Antibodies. A laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor.

Harpez, Y. & Chothia, C. (1994) Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains *J. Mol. Biol. 238*, 528-539.

Hawkins, R. E., Russell, S. J., Bay, M. & Winter, G. (1193) The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen. The interaction of mutant D1.3 antibodies with lysozyme. *J. Mol. Biol. 234*, 958-964.

Hendsch, Z. S., Jonsson, T., Sauer, R. T. & Tidor, B. (1996) *Biochemistry 35*, 7621-7625.

Hendsch, Z. S. & Tidor, B. (1994) *Protein Sci. 3*, 211-226.

Holliger, P. *et al.*, (1993) *Proc. Natl. Acad. Sci. 90*, 6444-6448.

Hu, S-z., *et al.*, *Cancer Res.* 56:3055-3061 (1996).

Ikura, M. & Bax, A. (1992) Isotope-filtered 2D NMR of a protein-peptide complex: study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc. 114*, 2433-2440.

Ikura, M., Kay, L. E. & Bax, A. (1991) Improved three-dimensional 1H-13C-1H correlation spectroscopy of a 13C-labeled protein using constant-time evolution. *J. Biomol. NMR 1*, 299-304.

Jacobs, J. & Schultz, P. G. (1987) Catalytic antibodies. *J. Am. Chem. Soc. 109*, 2174-2176.

Johnson, B. A. & Blevins, R. A. (1994) *J. Biomol. NMR 4*, 603-614.

Janda, K. D., *et al.*, *Science* 275:945-948 (1997).

Jones, E. Y. (1993) The immunoglobulin superfamily *Curr. Opinion struct. Biol. 3*, 846-852.

Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986)

Replacing the complementarity-determining regions in a human antibody with those from a mouse *Nature 321*, 522-525.

Kabsch, W. & Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577-2637.

Kamtekar, S. Schiffer JM, Xiong H, Babik JM, Hecht MH. (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262(5140):1680-1685.

Kauzmann, W. (1959) *Adv. Prot. Chem. 14*, 1-63.

Kay, L. E. (1995) Field gradient techniques in NMR spectroscopy. *Curr. Opinion Struct. Biol. 5*, 674-681.

Kay, L. E., Keifer, P. & Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity *J. Am. Chem. Soc. 114*, 10663-10665.

Kay, L. E. (1993) *J. Am. Chem. Soc. 115*, 2055-2057.

Kay, L. E., Xu, G.-Y. & Singer, A. U. (1993) A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations in H2O Samples of Proteins. *J. Magn. Reson. B 101*, 333-337.

Koide, S., Bu, Z., Risal, D., Pham, T.-N., Nakagawa, T., Tamura, A. & Engelman, D. M. (1999) *Biochemistry 38*, 4757-4767.

Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998) *J. Mol. Biol. 284*, 1141-1151.

Koide, S., Dyson, H. J. & Wright, P. E. (1993) Characterization of a folding intermediate of apoplastcyanin trapped by proline isomerization. *Biochemistry 32*, 12299-12310.

Kornblihtt, A. R., Umezawa, K., Vibe-Pederson, K. & Baralle, F. E. (1985) Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene *EMBO J. 4*, 1755-1759.

Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and scnematic plots of protein structures. *J. Appl. Cryst. 24*, 946-950.

Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. (1999) *Biochemistry 38*, 4896-4903.

Kunkel, T. A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA 82*, 488-492.

Leahy, D. J., Aukhil, I. & Erickson, H. P. (1996) *Cell 84*, 155-164.

Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. (1992) Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionlyl protein *Science 258*, 987-991.

Lee, W., Revington, M. J., Arrowsmith, C. & Kay, L. E. (1994) A pulsed field gradient isotope-filtered 3D 13C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS lett. 350*, 87-90.

Lerner, R. A. & Barbas III, C. F., *Acta Chemica Scandinavica*, 50 672-678 (1996).

Lesk, A. M. & Tramontano, A. (1992) Antibody structure and structural predictions useful in guiding antibody engineering. In *Antibody engineering. A practical guide.* (Borrebaeck, C. A. K., Ed.) Vol. W. H. Freeman & Co., New York.

Li, B., Tom, J. Y., Oare, D., Yen, R., Fairbrother, W. J., Wells, J. A. & Cunningham, B. C. (1995) Minimization of a polypeptide hormone *Science 270*, 1657-1660.

Litvinovich, S. V., Novokhatny, V. V., Brew, S. A. & Inhgam, K. C. (1992) Reversible unfolding of an isolated heparin and DNA binding fragment, the first type III module from fibronectin. *Biochim. Biophys. Acta 1119*, 57-62.

Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1992) Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS lett. 314*, 413-418.

Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999) *Biochemistry 38*, 16419-16423.

Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992) The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell 71*, 671-678.

Malakauskas, S. M. & Mayo, S. L. (1998) *Nat Struct Biol 5*, 470-475.

Masat, L., *et al.*, (1994) *PNAS* 91:893-896.

Martin, F., Toniatti, C., Ciliberto, G., Cortese, R. & Sollazzo, M. (1994) The affinity-selection of a minibody polypeptide inhibitor of human interleukin-6. *EMBO J. 13*, 5303-5309.

Martin, M. T., *Drug Discov. Today*, 1:239-247 (1996).

McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature 348,* 552-554.

McConnell, S. J., & Hoess, R. H., *J.Mol. Biol.* 250:460-470 (1995).

McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Koerner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. & Withers, S. G. (1996) *Biochemistry 35*, 9958-9966.

Metzler, W. J., Leiting, B., Pryor, K., Mueller, L. & Farmer, B. T. I. (1996) The three-dimensional solution structure of the SH2 domain from p55blk kinase. *Biochemistry 35*, 6201-6211.

Minor, D. L. J. & Kim, P. S. (1994) Measurement of the β-sheet-forming propensities of amino acids. *Nature 367*, 660-663.

Muhandiram, D. R., Xu, G. Y. & Kay, L. E. (1993) An enhanced-sensitivity pure absorption gradient 4D 15N,13C-edited NOESY experiment. *J. Biomol. NMR 3*, 463-470.

Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. (1995) Structure of the NH-kB p50 homodimer bound to DNA. *Nature 373*, 311-117.

Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995) *Protein Sci. 4*, 2138-2148.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS lett. 229*, 317-324.

Nilges, M., Kuszewski, J. & Brünger, A. T. (1991) in *Computational aspects of the study of biological macromolecules by nuclear magnetic resonance spectroscopy.* (Hoch, J. C., Poulsen, F. M. and Redfield, C., Ed.) Vol. pp. 451-455, Plenum Press, New York.

O'Neil *et al.*, (1994) in *Techniques in Protein Chemistry V* (Crabb, L., ed.) pp.517-524, Academic Press, San Diego.

O'Neil, K. T. & Hoess, R. H. (1995) Phage display: protein engineering by directed evolution. *Curr. Opinion Struct. Biol. 5*, 443-449.

Pace, C. N.. & Scholtz, J.M.. (1997) Measuring the conformational stability of a protein. In *Protein structure. A practical approach* (Creighton, T. E., Ed.) Vol. pp. 299-321, IRL Press, Oxford.

Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996) *Faseb J 10*, 75-83.

Pace, C. N., Laurents, D. V. & Erickson, R. E. (1992) *Biochemistry 31*, 2728-2734.

Parmley, S. F. & Smith, G. P. (1988) Antibody-selectable filamentous fd phage vectors: affinity purification of target genes *Gene 73*, 305-318.

Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D. & Kay, L. E. (1994a) Simultaneous acquisition of 15N- and 13C-edited NOE spectra of proteins dissolved in H2O. *J. Magn. Reson. B 103*, 197-201.

Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E. & Forman-Kay, J. D. (1994b) Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-g1 complexed with a high affinity binding peptide. *Cell 77*, 461-472.

Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000) *Nat Struct Biol 7*, 380-383.

Perutz, M. F., Gronenborn, A. M., Clore, G. M., Fogg, J. H. & Shih, D. T. (1985) *J Mol Biol 183*, 491-498.

Pessi, A., Bianchi, E., Crameri, A., Venturini, S., Tramontano, A. & Sollazzo, M. (1993) A designed metal-binding protein with a novel fold. *Nature 362*, 3678-369.

Pierschbacher, M. D. & Ruoslahti, E. (1984) *Nature 309*, 30-33.

Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. (1996) *Proc. Natl. Acad. Sci. USA 93*, 10703-10706.

Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. (1997) *J. Mol. Biol. 270*, 763-770.

Rees, A. R., Staunton, D., Webster, D. M., Searle, S. J., Henry, A. H. & Pedersen,

J. T. (1994) Antibody design: beyond the natural limits. *Trends Biotechnol.* 12, 199-206.

Roberts *et al.*, (1992) *Proc. Natl. Acad. Sci. USA 89*, 2429-2433.

Rosenblum, J. S. & Barbas, C. F. I. (1995) in *Antobody Engineering* (Borrenbaeck, C. A. K., Ed.) Vol. pp89-116, Oxford University Press, Oxford.

Sali, D., Bycroft, M. & Fersht, A. R. (1991) *J. Mol. Biol. 220*, 779-788.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A laboratory manual, 2nd Ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor.

Sandhu, G. S., Aleff, R. A. & Kline, B. C. (1992) Dual asymmetric PCR: one-step construction of synthetic genes. *BioTech. 12*, 14-16.

Santoro, M. M. & Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl *a*-chymotrypsin using different denaturants *Biochemistry 27*, 8063-8068.

Smith, G. P. & Scott, J. K. (1993) Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol. 217*, 228-257.

Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science 228*, 1315-1317.

Smith, C. K., Munson, M. & Regan, L. (1995). Studying α-helix and β-sheet formation in small proteins. Techniques Prot. Chem., 6, 323-332.

Smith, C. K., Withka, J. M. & Regan, L. (1994) A thermodynamic scale for the b-sheet forming tendencies of the amino acids. *Biochemistry 33*, 5510-5517.

Smyth, M. L. & von Itzstein, M. (1994) Design and synthesis of a biologically active antibody mimic based on an antibody-antigen crystal structure. *J. Am. Chem. Soc. 116*, 2725-2733.

Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B. & Raleigh, D. P. (2000) *Biochemistry 39*, 872-879.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes *Methods Enzymol. 185*, 60-89.

Suzuki, H. (1994) Recent advances in abzyme studies. *J. Biochem. 115*, 623-628.

Tello, D., Goldbaum, F. A., Mariuzza, R. A., Ysern, X., Schwarz, F. P. & Poljak, R. J. (1993) Immunoglobulin superfamily interactions. *Biochem. Soc. Trans. 21*, 943-946.

Thomas, N. R. (1994) Hapten design for the generation of catalytic antibodies. *Appl. Biochem. Biotech. 47*, 345-372.

Timasheff, S. N. (1992) *Curr. Op. Struct. Biol. 2*, 35-39.

Verhoeyen, M., Milstein, C. & Winter, G. (1988) Reshaping human antibodies: Grafting an antilysozyme activity. *Science 239*, 1534-1536.

Venturini *et al.*, (1994) *Protein Peptide Letters 1*, 70-75.

Vuister, G. W. & Bax, A. (1992) Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling. *J. Magn. Reson. 98*, 428-435.

Vuister, G. W., Clore, G. M., Gronenborn, A. M., Powers, R., Garrett, D. S., Tschudin, R. & Bax, A. (1993) Increased resolution and improved spectral quality in four-dimensional 13C/13C-separated HMQC-NOESY-HMQC spectra using pulsed filed gradients. *J. Magn. Reson. B 101*, 210-213.

Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T. & Winter, G. (1989) Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli *Nature 341*, 554-546.

Webster, D. M., Henry, A. H. & Rees, A. R. (1994) Antibody-antigen interactions *Curr. Opinion Struct. Biol. 4*, 123-129.

Williams, A. F. & Barclay, A. N., *Ann. Rev. Immunol.* 6:381-405 (1988).

Wilson, I. A. & Stanfield, R. L. (1993) Antibody-antigen interactions. *Curr. Opinion Struct. Biol. 3*, 113-118.

Wilson, I. A. & Stanfield, R. L. (1994) Antibody-antigen interactions: new structures and new conformational changes *Curr. Opinion Struct Biol. 4*, 857-867.

Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994) Making antibodies by phage display technology*Annu. Rev. Immunol. 12*, 433-455.

Wiseman, T., Williston, S., Brandts, J. F. & Lin, L.-N. (1989) Rapid

measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biolchem. 179*, 131-137.

Wittenkind, M. & Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beba-carbon resonances in proteins *J. Magn. Reson. B 101*, 201-205.

Wu, T. T., Johnson, G. & Kabat, E. A. (1993) Length distribution of CDRH3 in antibodies *Proteins: Struct. Funct. Genet. 16*, 1-7.

Wüthrich, K. (1986) NMR of proteins and nucleic acids, John Wiley & Sons, New York.

Yamazaki, T., Forman-Kay, J. D. & Kay, L. E. (1993) Two-Dimensional NMR Experiments for Correlating 13C-beta and 1H-delta/epsilon Chemical Shifts of Aromatic Residues in 13C-Labeled Proteins via Scalar Couplings. *J. Am. Chem. Soc. 115*, 11054.

Yang, A.-S. & Honig, B. (1992) *Curr. Opin. Struct. Biol. 2*, 40-45.

Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994) Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR 4*, 845-858.